공학석사 학위논문

Development of Empirical Equations and Estimation Method of Transient Storage Model Parameters for Solute Transport in Rivers

하천 오염물질 혼합 해석을 위한 저장대 모형의

매개변수 산정법 및 경험식 개발

2019년 8월

서울대학교 대학원

건설환경공학부

노 효 섭

Abstract

Analyses of solute transport and retention mechanism are essential to manage water quality and river ecosystem. As reported by tracer injection studies that have been conducted to identify solute transport mechanism, concentration curves measured in natural stream have steep rising and long tail parts. This phenomenon is due to solute exchange process between transient storage zones and the main river stream. The transient storage model (TSM) is one of the most widely used models for describing solute transport in natural stream, taking transient storage exchange process into consideration. In order to use this model, calibration of four TSM parameters is necessary. Inverse modelling using measured breakthrough curves (BTCs) from tracer injection test is general method for TSM parameter calibration. However, it is not feasible to carry out performing tracer injection tests, for every parameter calibration. For that reasons, empirical formulae with hydraulic data, which is comparatively easier to obtain, have been proposed for the purpose of parameter estimation. This study presents two methods for TSM parameter estimation. At first, inverse modelling method employing global optimization framework, Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL), that incorporating famous evolutionary algorithms in water resource management field, was suggested. Second, TSM parameter empirical equations were derived adopting Multigene Genetic Programming (MGGP) based symbolic regression library GPTIPS and using Principal Components Regression (PCR). In terms of general performance, equations of this study were superior to published empirical equations.

Keywords: solute transport, Transient Storage Model (TSM), parameter estimation, Multigene

Genetic Programming (MGGP), Principal Components Regression (PCR), Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL),

Student number: 2017-29424

Table of Contents

List of Table

List of Figures

List of Symbol

Latin Uppercase

A_f	area of cross sectional area of main flow zone
A_s	area of cross sectional area of storage zone
A_{STS}	area of cross sectional area of surface transient storage
A_{HTS}	area of cross sectional area of hyporheic transient storage
C_{bc}	solute concentration at the upper boundary
C_f	concentration of free flow zone
C_{ff}	Chezy-type friction coefficient
C_{init}	initial background concentration of in-stream
C_s	concentration of storage zone
C_D	drag coefficient of plant
C_{HTS}	concentration of hyporheic transient storage
C_K	coefficient of Pedersen's equation
C_L	concentration of lateral inflow
C_{Sinit}	initial background concentration of storage area
C_{STS}	concentration of surface transient storage
D_a	apparent dispersion coefficient
D_g	mean grain diameter

DaI	Damkohler number
E_z	lateral mixing coefficient
EMP	evolutionary method performance metric
ES	exponent matrix in evolutionary polynomial regression
EX	user defined exponent vector in evolutionary polynomial regression
F	stagnation zone accounting factor
$\overline{F}$	mean objective function value before evolution in each complex
$\overline{F}_N$	mean objective function value after evolution in each complex
Fr	Froude number
$H(x)$	Heaviside step function of x;
H_g	grain height
$I_1(x)$	modified Bessel function of first order and first kind
J_s	solute exchange flux working as sink-source term of equation
K	one-dimensional dispersion coefficient
K_f	longitudinal dispersion coefficient of free flow zone
K_F	fractional dispersion coefficient
L	reach length
L_0	Distance from the tracer injection point for complete mixing
L_R	cascade/riffles reach length
L_{STS}	stream-wise length of surface transient storage

M	released mass
Q	volumetric discharge of stream
R	meander bend radius of curvature
$R_{\min}$	minimum radius of curvature
Re	Reynolds number
R^2	coefficient of determinant
R_i^2	i the variable's coefficient of determination
S	channel ground slope
S_g	spacing between roughness elements
S_n	channel sinuosity
SI	sensitivity index
T	residence time
T_T	random variable
T_W	residence time in the surface area
U	mean flow velocity
U_f	free flow zone velocity
U_a	apparent flow velocity
VIF	variance inflation factor
W	channel width
W_{MI}	mean width of the converged stream

W_{STS}	width of surface transient storage
X	input variable matrix in evolutionary polynomial regression
X_j	candidate inputs in evolutionary polynomial regression
$\hat{Y}$	predicted value of model
$Z^j_{N\times m}$	product of input variable matrix X

Latin Lowercase

a	frontal area per canopy volume
a_j	constant coefficient of each term in the evolutionary polynomial regression
b_i	coefficients of the multi-gene genetic programming
d_{stem}	diameter of stem
d_E	mean depth at the mixing layer interface
d_{MI}	mean depth of the converged stream
d_{STS}	mean depth of the surface transient storage
e_{var}	elasticity
f	friction factor
g	gravitational acceleration
$g(\tau)$	memory function

h	mean flow depth
k	parameter accounting for solute adsorption
k_P	power law exponent
n	number of injection points
q_L	average lateral inflow per unit bed area
$r(t;x)$	probability density function of residence time
t	time
x	direction along the stream
x_{BFS}	detachment distance downstream from the backward-facing step
x_{FFS}	detachment distance downstream from the forward-facing step
y_{FFS}	reattachment distance for the upstream streambank

Greek Uppercase

ΔS	mean spacing between individual plant stem
ΔU	velocity difference between the main channel and wake region velocity

Greek Lowercase

α	exchange rate of storage zone
α_i	relevant rate ($=q_L/h$)

α_{STS}	exchange rate of surface transient storage
α_{HTS}	exchange rate of hyporheic transient storage
β	confluence angle
γ	measured from the upstream bottom to the obstacle
$\delta(t)$	Dirac delta function
δ_e	mixing layer penetration depth into the canopy
ε	storage area ratio
$\phi_i(t)$	probability density function of trapping event in i the storage
η_S	parameter that accounts for storage zones
ν	kinematic viscosity of water
λ_f	biochemical reaction parameters in the main flow zone
λ_s	biochemical reaction parameters in the storage zone,
ρ	fluid density
ρ_{corr}	Pearson's correlation coefficient
θ	porosity of surface transient storage
τ	characteristic time scale of solute exchange between storage zone and flow zone
χ	the dead zone parameter which is equivalent to $\chi^2 = A_f / A_s$

Chapter 1. Introduction

1.1 Necessity and Background of Research

Precise prediction of solute transport in natural stream is essential to manage stream and river networks properly. Therefore, ecologists, hydrologists, and practitioners are interested in the fate and transport process of solutes such as nutrient spiraling and pollutant spill. For example, managing water quality problem is accorded to biochemical aspect. Harmful algal bloom, especially, is regarded as a serious threat to inland water quality manipulation (Dybas, 2005; Brooks et al., 2016; Bohrman and Strauss, 2018). Due to urbanization, industrialization, and agriculture, pollutants of concern are often nutrients like nitrate and phosphorus (O'connor et al., 2010; Bohrman and Strauss, 2018; TT Le et al., 2018). Moreover, Inflow of nutrients intensifies harmful algal bloom. Many researchers regard that transient storage features affect nutrient retention, accompanying nutrient uptake of macrophyte since transient storage affects flow structure increasing solute residence time. Vallet et al. (1996) referred that geomorphic properties may have effects on nutrient uptake by changing exposure time of water and solute to biochemically reactive substrates, indirectly. As exposure time of reactive substrates is increased biotic and abiotic uptake can be facilitated (Bohrman and Strauss, 2018). Some investigators found that transient storage is important to nutrient uptake, especially taking account for hyporheic transient storage. By experimental studies, researchers found that transient storage is responsible for phosphorous and nitrate retention (Mulholland et al., 1997, Thomas et al., 2003). Similarly, other researchers concluded that change of natural and artificial structure because of restoration can change transient storage features as Figure 1.1, and it influences nutrient uptake (Ensign and Doyle, 2005; Rana et al., 2017; TT Le et al., 2018;

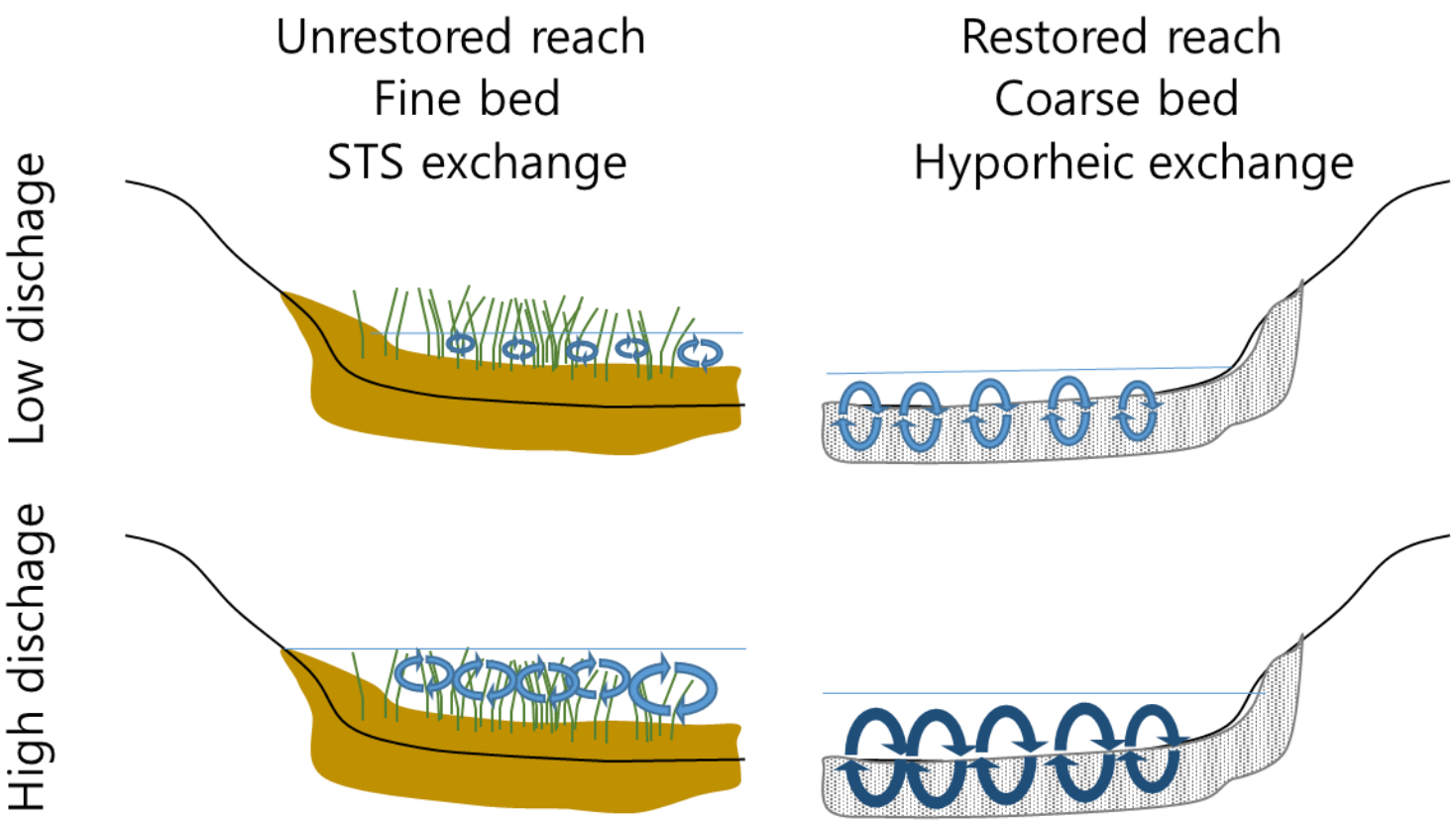


Figure 1.1 The schematic of mechanisms in the restored and unrestored reaches varying systematically with discharge. Arrows indicate hypothesized exchange. (adapted from Ward et al., 2018)

Ward et al., 2018). In addition, many researchers are working on identifying the source location of contamination in river system. In pollutant source identification in river system, hydrodynamic models, such as advection-dispersion model and transient storage model, are employed to generate data in order to consider complexity of river (Boano et al., 2005; Ghane et al., 2016; Zhang and Xin, 2017). Usually, pollutant source identification studies are organized as numerical model and statistical technique such as genetic algorithm, artificial neural networks. In terms of a geostatistical approach, Boano et al. (2005) used conventional TSM to generate data. Ghane et al. (2016) applied backward probability method, based on classical advection-diffusion equation and Saint-Venant equations, to identify the location and release time of pollutant solute.

In order to identify the process of solute transport, assessment and measurement of hydraulic, geomorphic, and concentration distribution curves are needed. In the situation without the needed data, spatiotemporal simulation of concentration is conducted. Many of researches about simulating the fate of solute were carried out using one-dimensional dispersion model. In natural stream, transient storages, as called as dead zone or stagnant zone, are varying along with the flow, e.g. bed morphology such as bed material, pool-riffle, channel meander, hydraulic structures, and aquatic vegetation. Solute transport is influenced by transient storage mass exchange mechanisms, resulting in different shape of concentration distribution data of tracer test. In high order natural stream, which has complex hydromorphic features, a number of tracer test data showed anomalous dispersion characteristics. Concentration data influenced by transient storage has peak with early detection, skewed distribution, longer tail at descending part, and non-linearly increasing variance of concentration distribution. In this manner, early studies have recommended models which consider storage effect for high order stream, rather than 1-D advection-dispersion model (Fischer, 1966; Hays, 1967; Day, 1975; Petersen, 1977; Beltaos and Day, 1978; Sabol and

Nordin, 1978; Liu and Cheng, 1980; Chatwin, 1980; Beer and Young, 1983, Bencala and Walters, 1983). To quantify transient storage effect, many investigators have been working on developing physical based models (Bencala and Walters, 1983; Haggerty et al., 2000; Wörman et al., 2002; Deng et al., 2004; Boano et al., 2007; Marion and Zaramella, 2005). The transient storage account models describe transient storage with additional parameters or terms. In most cases, those parameters are obtained from inverse modeling. However, extensive field tracer data is necessary to obtain the parameters by inverse modeling. Among the proposed phenomenological models, the Transient Storage Model (TSM) is the most frequently used model accounting for transient storage effect. Bencala and walters (1983) presented the governing equation of TSM as Eq. (1.1).

$$\frac{\partial C_f}{\partial t} = -\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f) + \alpha(C_S - C_f) \tag{1.1}$$

$$\frac{dC_s}{dt} = \alpha \frac{A_f}{A_s}(C_f - C_s) \tag{1.2}$$

where, t is the time; x is the direction along the stream; C_f, C_s, C_L are the concentration of free flow zone, storage zone, lateral inflow, respectively $[\mathrm{M}^3/T]$; Q is the volumetric discharge of stream $[\mathrm{L}^3/T]$; K_f is the longitudinal dispersion coefficient of free flow zone $[\mathrm{L}^2/T]$; A_f, A_s are the area of cross-sectional area of main flow zone and storage zone, respectively $[\mathrm{L}^2]$; α is the exchange rate of storage zone $[1/\mathrm{T}]$. The TSM simplifies river complexity partitioning river flow as two zones of the main free flow zone and storage zone

mass exchange rate coefficient, and main flow zone dispersion coefficient illustrated as Figure 1.2.

Bencala et al. (2011) pointed out that the simplicity of TSM is both its greatest strength and weakness in the meantime. Due to the simplicity, TSM is the most widely used model while it gives too simplified vision despite very complex environment. In other words, simplified four parameter leaves uncertain the relative importance of transient storage mechanism (Choi et al., 2000; Harvey et al., 2005; Wörman et al., 2007; Marion et al., 2008; Briggs et al., 2009). Taking most popular TSM model OTIS-P for instance, since OTIS-P exploits local search algorithm that is sensitive to the initial optimization parameter set, the best-fit parameter set can be a local minimum solution, instead of global solution (Kelleher et al., 2013). In some cases, the best-fitted parameter set is not reasonable (Scott et al., 2003, Ward et al., 2017). Mostly, this equifinality and local optimal solution problems are expressed as uncertainty in parameter estimation. A number of studies about assessment of TSM parameter estimation uncertainty (Runkel, 1998; Zaramella et al., 2016; Ward et al., 2017) in order to overcome uncertainty issues. By analyzing uncertainty of TSM parameters, many investigators are trying to solve uncertainty problem and figure out which factors contribute to uncertainty, e.g. Runkel (1998) employed uncertainty estimation into OTIS-P model. Wagner and Harvey (1997) and Davis and Atkinson (2000) noted that uncertainties of parameters are related with Damkohler number (DaI) accounting for, stream length, exchange rate, and stream water velocity. Ward et al. (2017) developed software tool OTIS-MCAT to assess uncertainty employing Monte-Carlo analysis into OTIS. Though the TSM has weakness of such uncertainty, investigators focus on correlating TSM parameters with measured physical values (e.g. Briggs et al. 2009; O'Connor et al., 2009; Stonedahl et al., 2013), since TSM parameter requires less computation cost and much simpler compared to other methods (Boano et al., 2014).

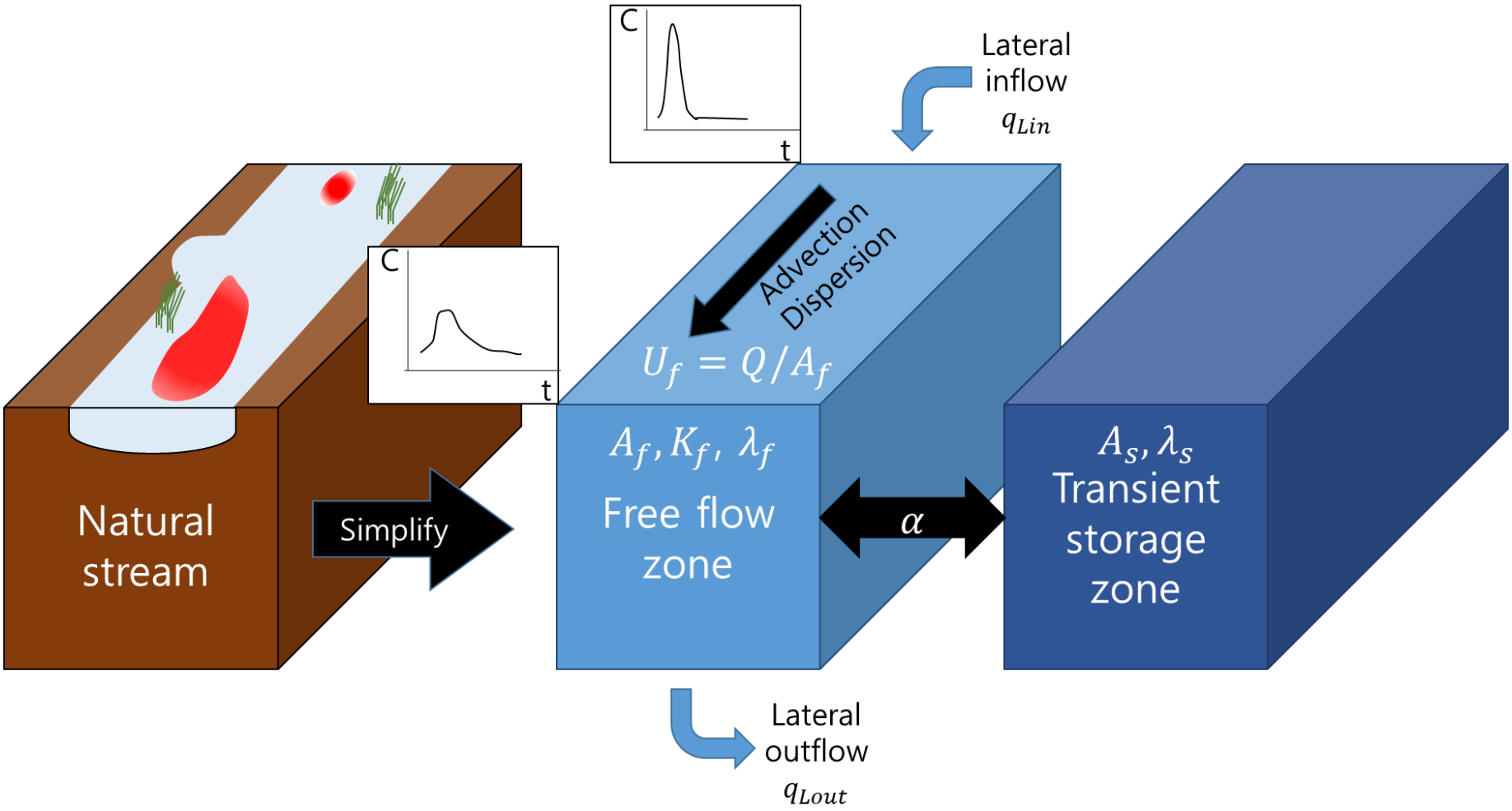


Figure 1.2 The schematic of Transient storage zone model

Efforts have taken in order to overcome the limitation of large-scale simulation that represented as necessity of extensive parameter calibration based on the field measurement. Empirical equations for predicting TSM parameters were developed to simulate without concentration data and to identify transient storage mechanism (Thackston and Schenelle, 1970; Pedersen, 1977; Bencala and Walters, 1983; Seo and Yu, 2000; Cheong and Seo, 2003; Cheong et al., 2007; Sahay, 2012; Femeena et al., 2019). Early works were focused on predicting storage zone area ratio, on the contrary, recent studies (Cheong and Seo, 2003; Cheong et al., 2007; Sahay, 2012; Femeena et al., 2019) were performed to identify relationship between TSM parameters (residence time T, flow zone dispersion coefficient K_f, and area ratio ε) with hydraulic variables. According to the published articles about TSM parameters and hydraulic feature, friction factors have a non-linear relationship with the TSM parameters. One-dimensional dispersion coefficient has been developed regarding flow width, flow depth, velocity, and shear velocity as important river characteristic (Taylor, 1954, Elder, 1959, Fischer, 1975, Seo and Cheong, 1998, Kashefipour and Falconer, 2002, Deng et al., 2001, McQuivey and Keefer, 1974, Disley et al., 2015, Alizadeh et al., 2017). Because the flow zone dispersion coefficient is similar to the one-dimensional dispersion coefficient, it is guessed that those stream characteristics play a role in TSM as well. Cheong et al. (2007) and Sahay (2012) developed equations for predicting TSM parameters adopting input variable Peclet number (= UL/K; where U is the stream mean flow velocity; L is the reach length; K is the one-dimensional dispersion coefficient) noting the importance of reach length and velocity. The equations were derived based on the pre-estimated one-dimensional dispersion coefficient from a study of Seo and Cheong (2003) by Routing method. The Routing method requires concentration data obtained from tracer test. In other words, breakthrough curves (BTCs) are necessary in order to use the equations using one-dimensional dispersion coefficient concentration. However, if the concentration data set and the hydromorphic data set are

measured, direct TSM parameter estimation is much desirable for having proper values. Certainly, user can use prediction equation to use one-dimensional dispersion coefficient, without direct estimation of dispersion coefficient, however, due to the uncertainty of the prediction equation itself, uncertainty can be propagated causing an unsatisfactory result. Thus, new prediction equation accounting for reach length effect instead of using Peclet number can be useful for the user who does not have concentration data. While on the subject though lots of researches were published for development of empirical equations for TSM parameters, only one study for empirical equation based on OTIS model exists. Femeena et al. (2019) proposed simple regression model for TSM parameters simply using discharge, channel width, and channel depth except for main flow zone area because they regarded the cross-sectional area of the stream as a known measured parameter.

The empirical equations for TSM were derived based on the classical regression techniques. Recently, advanced regression techniques have been applied to the derivation of equations. Cheong et al. (2007) used Weighted Robust Minimum Covariance Determinant Method. A Genetic Algorithm (GA) was exploited for the derivation of equations with Cheong et al. (2007)'s datasets (Sahay, 2012). However, whereas the fact that nonlinear relationship between TSM parameters and hydromorphic features was reported, these equations have derived under the assumption of certain formula structure. In this manner, symbolic regression technique can be an option to describe the nonlinear behavioral model.

Genetic Programming (GP) is a specialized GA technique for symbolic regression proposed by Koza (1992). GP benchmarked natural selection concept of GA (crossover, mutation, and selection mechanism) into finding the structure of the model. This data-driven technique is widely applied to the civil and environmental engineering field. Alavi and Gandomi (2012) presented a soil liquefaction model using linear GP. Multi-Gene Genetic Programming (MGGP)

is an advanced version of GP that combines standard GP. The MGGP also have been utilized in civil and environmental problems such as concrete creep formulation (Gandomi et al., 2016) and daily streamflow prediction (Mehr and Kahya, 2017). According to comparison study, in case of high dimension and large training set, GP was very competitive compared to Support Vector Machine (SVM) and Particle Swarm Model Selection (PSMS) (Valencia-Ram´ırez et al., 2014). Even though Artificial Neural Networks (ANNs) and SVMs showed successful prediction result in many of studies, ANNs and SVMs have problems of the potential for over-fitting, the challenge of kernel parameter tuning, and not transparent results (Gandomi et al., 2016). As remarked above, much stable and explicit empirical equations can be derived by using MGGP than other techniques (i.e. classical regression methods), for illustrating complexity of river system.

In summary, the transient storage model is used by ecologists, engineers, and practitioners to manage water quality in river system. In this manner, TSM parameters are used for understanding solute transport. By all means, estimating TSM parameters directly using tracer test data set is the best way to obtain TSM parameters. However, in real stream, it is difficult to perform a tracer injection test whenever parameter estimation is needed. Moreover, usually, tracer test data is scarce. Thus, the formulation of TSM parameters prediction can be used in many cases. For example, prediction formulae can be useful when identifying the location of solute is needed in a wide range of watershed or river reaches for which the results of tracer tests are not available, as no better technique for determining transient storage model parameters in such cases is known. As long as multiple zone storage models that estimate storage parameters are adopting exhausting methods for parameter estimation (Kerr et al. 2012; Ward et al., 2017), it is important to set small range of parameter space as possible. In such a manner, one-zone transient storage parameter equations can help for inserting reasonable parameter search space. Then as well, the formulae can help to approximate initial parameter

set formulae, for who want to use OTIS-P. Moreover, for the purpose of determination of experimental reach length of tracer test taking Wagner and Harvey (1997)'s remark into reference, Damkohler number can be calculated with provided initial TSM parameter set. Considering each developed formula has explicit structure, a parameter sensitivity analysis was conducted. In addition, developed formulae were coupled with the pseudo numerical model of OTIS that is implemented to TSM parameter estimation framework, for alternative use, so that users can simulate TSM without calibration.

Moreover, TSM parameter obtaining sequence is suggested as Figure 1.3 for users who need TSM parameter. The first requirement for obtaining TSM parameter is hydraulic and geometric data because proposed equations and parameter estimation frameworks require hydraulic data. Then the user should determine the parameter obtaining method under the condition of existing concentration distribution data.

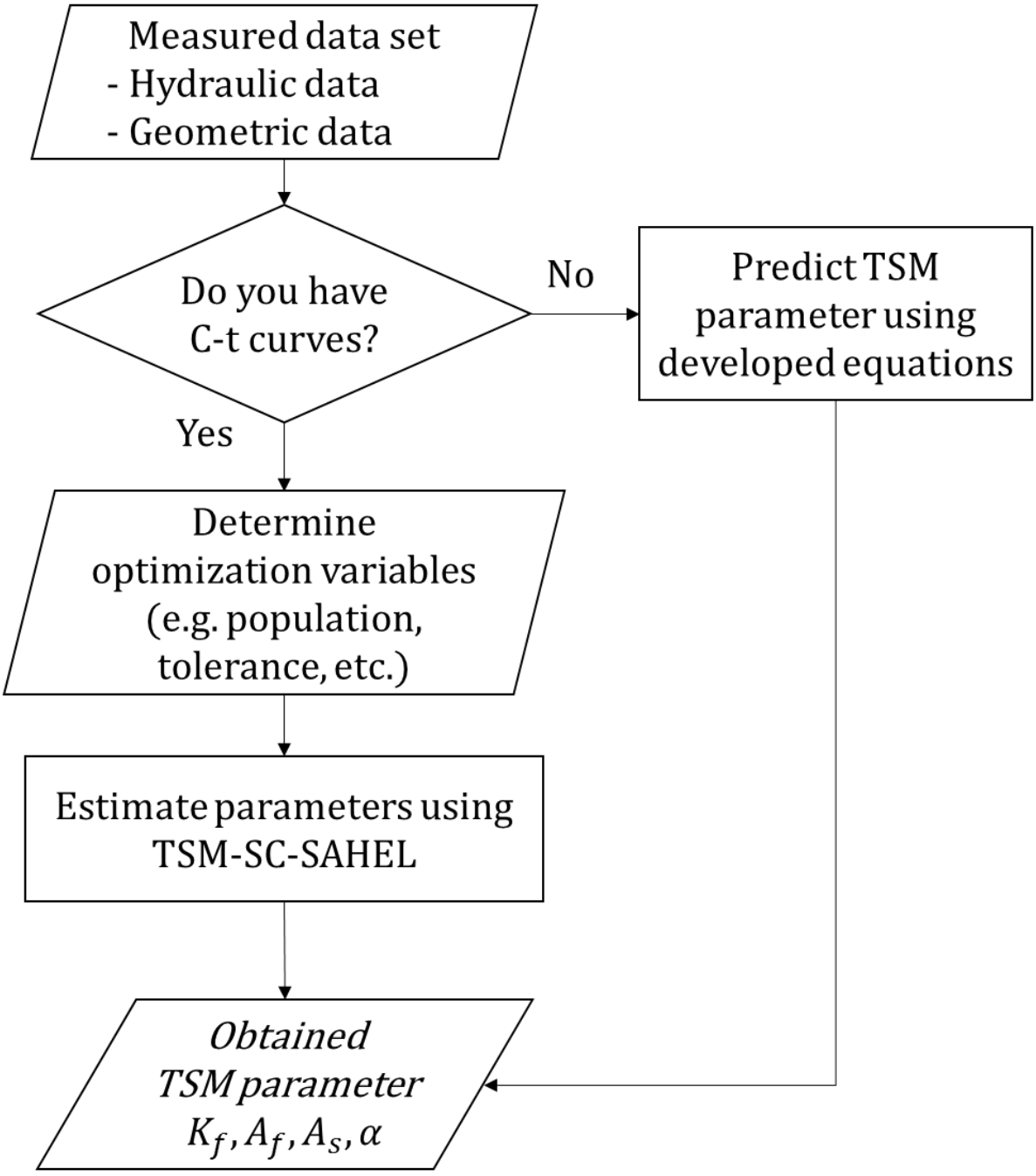


Figure 1.3 The flowchart of obtaining TSM parameter

1.2 Objectives

The main objective of this study is to propose parameter estimation framework and new formulae for TSM parameters for non-Fickian solute transport in natural streams. For development of formulae, multi-gene genetic programming method was applied. In addition, TSM-based parameter estimation framework, that estimates TSM parameters from concentration curves, was developed using meta-heuristic optimization technique. In order to select input variables, dimensional analysis was conducted. Thereafter, equations for each TSM parameter were derived by applying Multi-Gene Genetic Programming (MGGP) method. The detailed methodology of this study is illustrated in Figure 1.4, and given below.

(1) Tracer test data sets were collected from previous field tracer experiment studies in order to apply parameter estimation framework to the concentration curves. The framework contains TSM numerical model accorded with meta-heuristic adaptive optimization algorithm SC-SAHEL (Shuffled Complex-Self Adaptive Hybrid EvoLution). Numerical scheme of the numerical model is a Crank-Nicolson scheme which is adopted on OTIS (Runkel, 1998). To guarantee reliability of TSM parameters, mesh independency test was conducted. Using presented parameter estimation framework, TSM-SC-SAHEL, TSM parameters were estimated from the collected data set. In addition to direct estimation, estimated TSM parameter sets using OTIS-P form published literatures were collected. Thus, both published OTIS based parameter sets and directly estimated parameters were included in model derivation dataset.

(2) Since case specificity and uncertainty have been considered as a problem by TSM investigators in TSM parameter estimation (Gooseff, 2013; Kelleher, 2013), several numerical experiments were conducted for accuracy and efficiency of parameter estimation.

As pointed out by Gooseff (2013), difference of estimated parameter values on same stream in different section combination was confirmed. Furthermore, in the numerical point of view, mesh independence of estimated TSM parameter was analyzed. To identify optimal setting of optimization, compatibility of objective functions and optimization methods were discussed. According to the assessments, recommendable setting and algorithm were adopted as default option of TSM parameter estimation framework presented in this study.

(3) For determining input variables of developed empirical equations, relationship of variables was analyzed. In order to select physically meaningful variables from datasets, previous researches about classification of transient storage solute exchange mechanism (Tonia and Buffington, 2009; Jackson et al., 2013) were taken into account. Dimensional analysis was conducted in order to make generalized formulae. As a result, relationships between dimensionless TSM parameter and dimensionless variables were evaluated.

(4) New empirical equations for TSM parameters were derived using MGGP which is a symbolic data-driven regression technique. During the derivation of equations, the structure of equation is searched by a Genetic Algorithm (GA). Then the Least Squares (LS) solution is obtained to find proper coefficient of equations. Throughout derivation of equation, the MGGP tries to find a simple and accurate model. In other words, the MGGP performs multi-objective optimization taking objective functions of low complexity and model fitness, and finds Pareto front of multi-objective optimization.

(5) The performances of equations for TSM parameters were evaluated. Using estimated TSM parameters in (1), predicted values from equations were compared. In addition, predicted BTCs using developed equations and observed BTCs are compared to judge the applicability of equations.Then, the weakness and advantage of previously published and developed equations from this study were assessed.

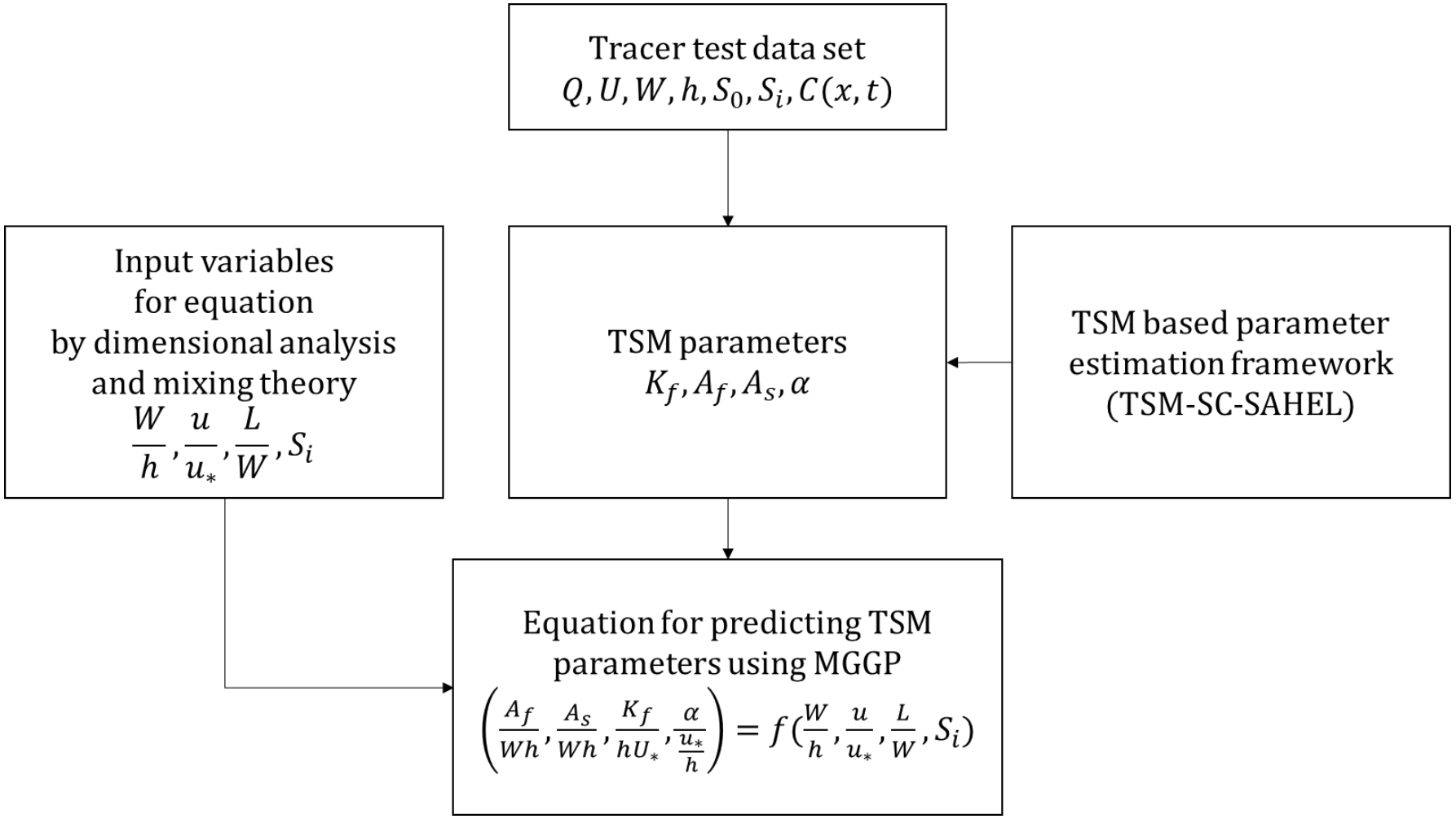


Figure 1.4 The flowchart of this study for prediction of transient storage zone model parameter

Chapter 2. Theoretical Background

2.1 Transient Storage Model

2.1.1. Mechanisms of Transient Storage

Solute transport is influenced by Mass exchange process between such storage zone and free flow zone. Transient storage is discerned as sum of surface transient storage (STS), which is in-stream flow structures and hyporheic transient storage (HTS), which is shallow groundwater flow beneath alluvial valley (Figure 2.1). Solutes are temporarily detained in the hyporheic zone and surface properties significantly changing velocity (Bencala and Walters, 1983; Harvey et al., 1996; Boulton et al., 1998; Briggs et al., 2009; Tonia and Buffington, 2009; Buffington and Tonia, 2009; Jackson et al., 2013).

Mixing in hyporheic transient storage referred to as hyporheic exchange in river occurs via circulation cells moving water into subsurface alluvium and back again (Vaux, 1962; Bencala, 2005; Tonia and Buffington, 2009; Buffington and Tonia, 2009). It is reported that the downwelling and upwelling mechanism is generated by pressure difference between downwelling point and upwelling point. As hyporheic exchange transfer solutes and surface water into the sediment and back again, solute concentration is changed (Bencala and Walters, 1983). For reactive solute with sediment organism, upwelling flow brings transformed solute to river and this can be a source of nutrients (Triska et al. 1989; Nagaoka and Ohgaki, 1990; Valett et al. 1994;).

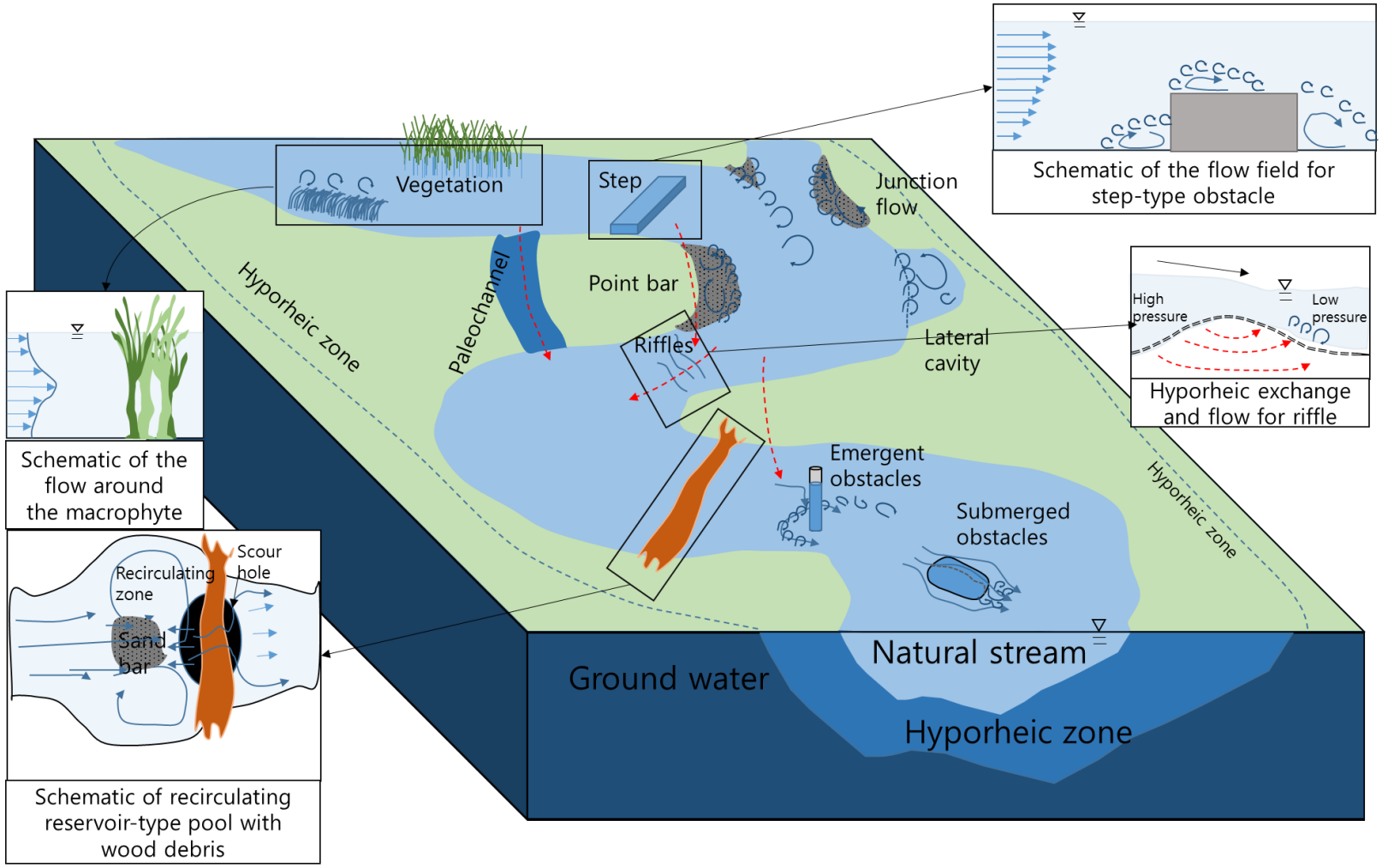


Figure 2.1 Schematics of surface, hyporheic transient storage in alluvial valley
(red dashed lines indicate hyporheic exchange and navy solid lines indicate free flow through surface transient storage)

Table 2.1 Classification of surface transient storage by properties

(Adapted from Jackson et al., 2013)

Type of STS	Properties
Lateral cavity	
Emergent	Propagated mixing layer in entrance and recirculation zone in cavity
Submerged	Main channel stream-wise overtopping flow
Bulging obstacles	
Backward-facing step	Recirculation zone behind obstacle
Forward-facing step	Recirculation zone in front of and behind obstacle
Isolated flow obstacles	
Emergent	Horseshoe vortex facing and von Karman vortex structure behind obstacle
Submerged	Closed recirculation zone behind obstacle
Cascades and riffles	Coalescence of wake fields (cylindrical shape)
Macrophyte	
Emergent	Coalescence of wake fields behind macrophytes
Submerged	Mixing layer top of vegetation and a monami phenomenon
Pools	
Vertically submerged cavity	Pool bathymetry sufficiently below upstream reach bathymetry
Closed lateral cavity	Lateral cavity flow field has a backward- and forward-facing step
Recirculating reservoir	*Jet-like flow*: flow enters and generates circulating flow zone *Flow impingement*: flow impinges, deflects, and recirculates *Scour pool*: vertical flow enters, scours pool head, and recirculates
Meander bends	Inner bank mixing later forms recirculation zone nad point bar and outer bank mixing layer scours and erodes stream bank
Junction flow	Velocity differences between confluencing stream governing interface position, mode, and coherence

Table 2.2 Hydromorphic parameters about classified surface transient storage types (adapted from Jackson et al., 2013)

Type of STS	Flow	Roughness	Constants	Case-specific parameters
Lateral cavity				
Emergent	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, L_{STS}$
Submerged	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, L_{STS}, (h-d_{STS})$
Bulging obstacles				
Backward-facing step	U	U_*	g ν	$d_{E,BFS}$, $d_{STS,BFS}$, W_{STS}, x_{BFS}, γ
Forward-facing step	U	U_*	g ν	$d_{E,BFS}, d_{E,FFS}$, $d_{STS,BFS}, d_{STS,FFS}$, W_{STS}, $x_{BFS}, x_{FFS}, y_{FFS}$, γ
Isolated flow obstacles				
Emergent	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, W, \theta$
Submerged	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, W, \theta, (h-d_{STS})$
Cascades and riffles	U	U_*		$h, D_g, H_g, L_R, S_g, (h-H_g)$
Aquatic vegetation				
Emergent	U	C_D	g	$a, d_{stem}, \Delta S$
Submerged	U	C_D	g	$a, d_{stem}, S, h, d_C, \delta_e, \Delta U$
Pools				
Vertically submerged cavity	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, L_{STS}$
Closed lateral cavity	U	U_*	g ν	$d_{E,BFS}, d_{E,FFS}$, $d_{STS,BFS}, d_{STS,FFS}$, W_{STS}, $x_{BFS}, x_{FFS}, L_{STS}$,
Recirculating reservoir	U	U_*	g ν	$d_E, d_{STS}, W_{STS}, L_{STS}, \theta$
Meander bends	U	C_f	g ν	$d_E, d_{STS}, W_{STS}, L_{STS}, W, R, R_{\min}$
Junction flow	U_1, U_2	C_f	g ν	d_E, d_{STS}, W_{STS}, L_{STS}, d_{MI}, W_{MI}, $W, R, R_{\min}, \beta$

The values used in this table are the following: g is the gravitational acceleration $[\mathrm{L}/\mathrm{T}^2]$; ν is the kinematic viscosity of water $[\mathrm{L}^2/\mathrm{T}]$; W_{STS}, and L_{STS} are the width and stream-wise length of STS, respectively $[\mathrm{L}]$; d_{STS} is the mean depth of the STS $[\mathrm{L}]$; d_E is the mean depth at the mixing layer interface $[\mathrm{L}]$; $(h-d_{STS})$ is the submerged level $[\mathrm{L}]$; h is the mean flow depth $[\mathrm{L}]$; x_{BFS} is the detachment distance downstream from the backward-facing step $[\mathrm{L}]$; γ is the angle measured from the upstream bottom to the obstacle; x_{FFS} is the detachment distance downstream from the forward-facing step $[\mathrm{L}]$; y_{FFS} is the reattachment distance for the upstream streambank $[\mathrm{L}]$; W is the channel width $[\mathrm{L}]$; θ is the porosity of STS obstacles; D_g is the mean grain diameter $[\mathrm{L}]$; H_g is the grain height $[\mathrm{L}]$; L_R is the cascade/riffles reach length $[\mathrm{L}]$; S_g is the spacing between roughness elements $[\mathrm{L}]$; a is the frontal area per canopy volume $[1/\mathrm{L}]$; d_{stem} is the diameter of stem $[\mathrm{L}]$; ΔS is the mean spacing between individual plant stem $[\mathrm{L}]$; S is the channel slope; δ_e is the mixing layer penetration depth into the canopy $[\mathrm{L}]$; ΔU is the velocity difference between the main channel and wake region velocity $[\mathrm{L}/\mathrm{T}]$; C_D is the drag coefficient of plant; C_{ff} is the Chezy-type friction coefficient; R is the meander bend radius of curvature $[\mathrm{L}]$; $R_{\min}$ is the minimum radius of curvature $[\mathrm{L}]$; W_{MI} is the mean width of the converged stream $[\mathrm{L}]$; d_{MI} is the mean depth of the converged stream $[\mathrm{L}]$; β is the confluence angle.

Most of surface transient storages affect solute transport with two hydrodynamic factors of wakes and recirculating zone. These features are described as free shear flow structure, e.g. jets, wakes, and mixing layers (Versteeg and Malalasekera, 2007). Free shear flows consist of coherent structures recognized as disordered flow (Socolofsky and Jirka, 2005; Versteeg and Malalasekera, 2007). These coherent structures are larger scale turbulence formed

in the region of velocity shear. The velocity shear region is formed by instabilities in which smaller vortices interact by mechanisms like paring, tearing, or stretching resulting in larger scale vortex structures (Socolofsky and Jirka, 2005). Vortices are advected in stream, constructing and destroying each other. Some STS type has the coalescence of similar free shear flows or differing free shear flow (Jackson et al., 2013). Jackson et al. (2013) classified eight types of surface transient storage and accounting hydromorphic parameters based on differing mean flow properties as Table 2.1 – 2.2. The classification of STS consist of following categories: (1) lateral cavities (emergent and submerged); (2) bulging obstacles (backward- and forward-facing step); (3) isolated flow obstacles (emergent and submerged); (4) cascades and riffles; (5) macrophyte (emergent and submerged); (6) pools (vertically submerged cavity, closed cavity, and recirculating reservoir); (7) meander bends; and (8) Junction flow.

2.1.2. Models Accounting for Transient Storage

The advection-dispersion equation (ADE) (Eq. 2.1) has been standard model for solute transport simulation since 1970s (Fischer et al, 1979; Boano et al., 2014). The mass conservation equation of ADE is given as

$$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} = K\frac{\partial^2 C}{\partial x^2} \tag{2.1}$$

where, C is the solute concentration transverse averaged $[\mathrm{M}^3/\mathrm{T}]$. The ADE follows assumption of Brownian motion so-called Fickian advection diffusion process. The early investigators of dispersion pointed that the limitation of ADE that cannot fully describes observed concentration distribution data (Thackston and Schnelle, 1970; Nordin and Sabol, 1974; Valentine and Wood, 1977; Fischer et al., 1979; Nordin and Troutman, 1980). Bencala and walters (1983) analyzed Uvas Creek experiments and explained anomalous dispersion process with surface transient storage referred as “dead zones”. Due to such limitation of ADE, especially in natural stream tracer injection experiments, many efforts have focused on extending and generalizing ADE in recent two decades as summarized in Table 2.3. Presented models are generalized or extended version of ADE and TSM. It implies extended models can express transient storage effects much precisely than conventional ADE.

The TSM (Bencala and Walters, 1983) is simplified model accounting transient storage effect as four parameterizations. Basically, TSM is based on assumption Fickian in-stream transport and first order mass transfer. To distinguish different transient storages, the multiple-

zone transient storage model (nZ-TSM) was developed (Choi et al., 2000) adding two more parameters for each added storage zone. Kerr et al. (2013) coded 2Z-TSM in Matlab (OTIS-M) and observed difference in which parameters are estimated under the condition of independent storages and Linked storages, respectively. Although multiple-zone TSM can differ transient storage mechanisms, it also has weakness of uncertainty in structure of transient storage.

The Fractional advection-dispersion equation model (FADE) is classified within space fractional and time fractional advection equation model. The FADE is modeled taking derivatives at time or space term of ADE. In this study, Space FADE (Deng et al, 2004) is presented at 2.1.2.5. The FADE produces asymptotic outcome of motion which is discontinuous in time or space. The FADE describes Lévy motion or subordinate motion showing anomalous diffusion process. The calibration parameters are velocity, dispersion coefficient, and order of fractional term. Direct measurement in hyporheic zone is recommended for the estimation of Time-FADE and TS-FADE. Moreover, since fractional orders mean anisotropic extent of the medium caused by storage zone, as simulation reach length increases, the fractional factor can be changed in some cases (Singh, 2008).

The Modified advection-dispersion model (MADE) proposed by Singh (2003) is constructed with additionally adopted three calibration factors. Singh (2008) conducted a comparison study of TSM, Space-FADE, and MADE. According to the study, the MADE requires less computation cost than other presented models because it uses an analytical solution (Singh, 2008). However, parameters of the MADE are still lumped parameter, and spatial differences cannot be applied. Following four models are modeled by adding stochastic sink source term to conventional ADE. In the models, the storage process of models is defined with memory function or residence time distribution.

Haggerty et al. (2000) proposed The Multirate Mass Transfer Model (MRMT). The sink

source term of the MRMT is defined as convolution product of memory function. Power law function is assumed for the memory function, the parameter dimension is four as same as TSM.

The Advective Storage Path Model (ASP) was developed by Wörman et al. (2002). The ASP describes advective flow in the HTS. Sink source term of the ASP is defined as residence time distribution.

The continuous time random walk approach (CTRW) describes Brownian or Lévy motion by memory function defined sink source term. Boano et al. (2007) proposed a model applying Continuous time random walk approach to stream transport. The CTRW considers a sequence of random jumps with different lengths and duration with varying memory function. The CTRW with the decoupled formulation is very useful in most cases, however, it will not be valid, especially, with fast pore water flows (Boano et al., 2014). In addition, Boano et al. (2014) pointed out that despite the flexibility of formulation, still available data for independently parameterizing STS and HTS is not sufficient.

Marion and Zaramella (2005) proposed The Solute Transport in Rivers (STIR) in order to overcome the difficulty of separating individual processes of STS and HTS. The method is based on the assumption that individual storage trapping events are independent of other storage histories so that residence time distribution (RTD) can be convolution of each RTD. Since STIR takes into account the number of storages, it appears physically reasonable. Nevertheless, if a number of storages are considered, the parameter dimension may be high.

Davis and Atkinson (2000) derived an analytical solution of the Aggregated Dead Zone (ADZ) model. The ADZ is a simplified model of TSM which is under the assumptions of that turbulent dispersion process of free flow zone is not accountable for the fate of solutes compared to dead zone exchange process.

Boano et al. (2014) presented a more comprehensive review of transient storage exchange and presented phenomenological models.

In complex natural river system, considering individual transient storage effects is not feasible giving nonunique solution (Boano et al., 2014). Therefore, lumped simplified model, TSM, can be recommendable in reach scale simulation. Even though some studies pointed out that TSM cannot distinguish STS and HTS explicitly, many investigators are working on identifying physical and biochemical mechanism using TSM parameters as aforementioned in Chapter 1. The mechanism of one-zone transient storage model (1Z-TSM) is easy to understand intuitively and costs less computation resource compared to other presented methods. Since studies about understanding solute transport focus on not only simulation of solute fate but also understanding physical factors from estimated TSM parameters, prediction equations of 1Z-TSM parameters can be used in a wider range of research field than other phenomenological storage models. In addition, once TSM key parameters have been obtained, 1Z-TSM can simulate reactive solute transport without further calibration, since reactive terms are separated in model expression. However, other models have to calibrate the parameters unless reaction terms are added to the governing equation. Therefore prediction equations of 1Z-TSM are much meaningful than of other model's parameters. In this manner, 1Z-TSM introduced in 2.1.2.1 was adopted to estimate TSM parameter using tracer test data sets.

2.1.2.1 The one Zone Transient Storage Model (1Z-TSM)

Most widely used transport model taking into account storage zone effect is 1 zone transient storage model. The TSM proposed by Bencala and Walters (1983). The equations describe the advection, dispersion, and transient storage. The exchange process between the main channel

and storage zone is modeled as a first-order mass transfer as following equations. Spatiotemporal variation of solute including biochemical reaction term is formulated as follows.

$$\frac{\partial C_f}{\partial t} = -\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f) + \alpha(C_S - C_f) - \lambda_f C_f \quad (2.2)$$

$$\frac{dC_s}{dt} = \alpha \frac{A_f}{A_s}(C_f - C_s) - \lambda_s C_s \quad (2.3)$$

where, λ_f and λ_s are the biochemical reaction parameters in the main flow zone and storage zone, respectively.

2.1.2.2 The two Zone Transient Storage Model (2Z-TSM)

Since original TSM uses lumped parameter of one storage zone, describing storage processes that occur simultaneously in multiple storage zones, effects of surface transient storage and hyporheic transient storage were explained as one, although each of them has different storage mechanism. Therefore Choi et al. (2000) proposed 2-zone transient storage zone model and modified OTIS as 2-zone model.

$$\begin{aligned}\frac{\partial C_f}{\partial t} = &-\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f) \\ &+ \alpha_{STS}(C_{STS} - C_f) + \alpha_2(C_{HTS} - C_f) - \lambda_f C_f\end{aligned} \quad (2.4)$$

$$\frac{dC_{STS}}{dt} = \alpha_{STS} \frac{A_f}{A_{STS}} (C_f - C_{STS}) - \lambda_{STS} C_{STS} \quad (2.5)$$

$$\frac{dC_{HTS}}{dt} = \alpha_{HTS} \frac{A_f}{A_{HTS}} (C_f - C_{HTS}) - \lambda_{HTS} C_{HTS} \quad (2.6)$$

where, C_{STS} and C_{HTS} are the concentration of STS, HTS, respectively $[\mathrm{M}^3/T]$; A_{STS} and A_{HTS} are the area of the cross-sectional area of STS, HTS, respectively $[\mathrm{L}^2]$; α_{STS} and α_{HTS} are the exchange rate of STS, HTS, respectively $[1/T]$.

2.1.2.3 The Continuous Time Random Walk Approach (CTRW)

Application of fractional derivatives to Fickian theory is also widely applied to study about dispersion of heterogeneous media. Recently, the continuous time random walk approach (Montroll and Weiss, 1965; Scher and Lax, 1973) was applied to analysis solute transport under structure of heterogeneous and porous media. (Berkowitz and Scher, 1995). The continuous time random walk modeling approach for analysis of solute transport is written in the mathematical model as follows (Boano et al., 2007).

$$\frac{\partial C}{\partial t} + U \frac{\partial C}{\partial x} - K \frac{\partial^2 C}{\partial x^2} = J_s \quad (2.7)$$

where J_s is the solute exchange flux working as sink-source term of equation. This term is

related to the distribution of solute residence times in the subsurface, hyporheic zone, $f(T)$ and boundary based on Fick's law, $q = -K\nabla C$ (Elliott and Brooks, 1997; Boano et al., 2007). Unlike TSM that assumes distribution of exponential residence time distribution, CTRW approach allows any appropriate formulation of $f(T)$ and q (Boano et al., 2007).

2.1.2.4 The Modified Advection-Dispersion Model (MADE)

Singh (2003) proposed MADE model as follows.

$$K\frac{\partial^2 C}{\partial x^2} - U\frac{\partial C}{\partial x} = (1-\eta_S + k)\frac{\partial C}{\partial t} \tag{2.8}$$

where η_S is the parameter that accounts for storage zones; k is the parameter accounting for solute adsorption. The solution of MADE is given as

$$C = \frac{U_a}{2\sqrt{\pi K_a t}}\left(\frac{M}{Q}\right)_a \exp\left(-\frac{(x-U_a t)^2}{4K_a t}\right) \tag{2.9}$$

$$U_a = \frac{u}{(1-\eta+k)} \tag{2.10}$$

$$K_a = \frac{K}{(1-\eta+k)} \tag{2.11}$$

where, U_a is the apparent flow velocity $[\mathrm{L}/T]$; D_a is the apparent dispersion coefficient $[\mathrm{L}^2/T]$; M is the released mass $[\mathrm{M}]$. Optimization technique is used for estimating U_a, K_a, and $\left(M/Q\right)_a$. This model requires an observed concentration data in a single section. With estimated parameter, concentration can be calculated with solution equation (2.9). Hence, less computation time is needed compared to other models.

2.1.2.5 The Fractional Advection-Dispersion Equation Model (FADE)

The FADE model was presented by Deng et al. (2004). The model represents dead zone effect with higher order equations. The introduce equation (2.12) is special case of space fractional ADE. It is described by the following equations for numerical solution.

$$\frac{\partial(-J)}{\partial x} - U\frac{\partial C}{\partial x} = \frac{\partial C}{\partial t} \tag{2.12}$$

$$J = -K_F \frac{\partial^{F-1} C}{\partial x^{F-1}} \tag{2.13}$$

where, F is the stagnation zone accounting factor; K_F is the fractional dispersion coefficient $[\mathrm{L}^2/T]$. In which $F=2$ dispersion process is same as Fick's law. Adopting differential order F to governing equation of advection-dispersion model, it can express anisotropic, non-Fickian dispersion process. For isotropic media, F becomes value of 2. Deng et al. (2004) found that in most case $F=1.65$ in the range of 1.4 to 2.0. Since in most of stream with

storage zones, F is smaller than 2, numerical solution of FADE is comparatively complex.

2.1.2.6 The Multirate Mass Transfer Model (MRMT)

Haggerty et al. (2000). Governing equation is identical to Eq. (2.7). The mass exchange term with immobile domain is proposed in literature (Carrera et al., 1998) as Eq (2.14).

$$J_s = \frac{\partial C}{\partial t} g(t) = \int_0^t \frac{\partial C(x, t-\tau)}{\partial \tau} g(\tau) d\tau \tag{2.14}$$

where, $g(\tau)$ is the memory function. The memory function is integral function of exponential distributions weighted by power-law function as follows

$$g(t) = \beta_{tot} \int_{\alpha \min}^{\alpha \max} \frac{(k_P - 2)\alpha^{k_P - 2} e^{-\alpha t}}{\alpha_{\max}^{k_P - 2} - \alpha_{\min}^{k_P - 2}} d\alpha \tag{2.15}$$

where, β_{tot} is the capacity coefficient; $\alpha_{\max}^{k-2}$ and $\alpha_{\min}^{k-2}$ are maximum and minimum exchange rate, respectively $[1/T]$; k_P is the power law exponent. The model is understood as weighted combination of an infinite number of finite domain acting on timescale range of $\alpha_{\max}^{k-2}$ and $\alpha_{\min}^{k-2}$. Consequently, The MRMT has four calibration parameter having an equivalent degree of uncertainty as the TSM.

2.1.2.7 The Advective Storage Path Model (ASP)

Wörman et al. (2002) proposed the advective storage path model (ASP) as special case of MRMT. The ASP describes case of purely advective hyporheic exchange driven by pressure pumping. Hence ASP is appropriate model for purely advective hyporheic stream. Governing mass transport equation is same as MRMT and CTRW (2.7). The ASP is under assumption of Fickian diffusion or first-order diffusion in storage domain. The transport equation in hyporheic zone of the ASP model is as follows.

$$\frac{\partial \theta C_s}{\partial t} + \frac{1}{A}\frac{\partial A V_\eta C_s}{\partial \eta} = 0 \tag{2.16}$$

Where, η is curve-linear coordinate along hyporheic flow path; V_η is the hyporheic advection velocity $[\mathrm{L}/T]$. Subsurface velocity is assumed that constant along path line. Dissolved solute exchange is the product of exchange velocity perpendicular to surface and solute concentration in water. in the same manner, net concentration flux become as Eq (2.17)

$$J_s = \frac{1}{2}\int_0^\infty f(T)\frac{P}{A_f}\xi\left[-V_\eta(\tau,T)|_{\tau=0}\, C_f + (V_\eta(\tau,T)C_s)|_{\tau=T}\right]dT \tag{2.17}$$

where, ξ is the area reduction factor equivalent to ratio of normal velocity and hyporheic velocity V_n / V_η; T is the residence time $[T]$; P is the wetted perimeter of the stream$[\mathrm{L}]$.

2.1.2.8 The Solute Transport in Rivers Model (STIR)

Other approaches are focusing on spatiotemporal effect of hyporheic exchange separating individual process of STS and HTS. However, in some cases, HTS and STS are acting simultaneously so that difficult to distinguish two processes. Marion and Zaramella (2005), Marion et al. (2008) proposed the STIR model is probabilistic framework providing means to integrate each solution. The solute transport process along a stream is treated as stochastic process. The needed time to travel distance x is random variable T_T with probability density function (pdf) $r(t;x)$. The T_T is the sum of time in the surface area T_W ,with pdf $r_W(t;x)$ and spent time in storage area $T_S = \sum_{i=1}^{N} T_{Si}$, that is sum of time in single storage domains. A particle can move to i the storage and back to main flow zone independently. It is under the assumption of no longitudinal displacement in storage domain is negligible. The number of times trapped in i the storage domain is also random variable N_i with conditional distribution $p(n \mid T_W = t_W)$. The particle release time after trapped is determined with pdf $\phi_i(t)$. With the assumption of independent trapping event, spent time of n the trapping event in i the storage zone is as follows.

$$r_{Si|n}(t) = \phi_i(t) * \ldots * \phi_i(t) = \left[\phi_i(t)\right]^{*n} \tag{2.18}$$

where the symbol $(*)$ denotes time convolution, with $\phi_i(t) * \phi_i(t) = \int_0^t \phi_i(\tau)\phi_i(t-\tau)d\tau$. When number of trapping is equal to zero, $r_{Si|0}(t) = \delta(t)$ that is Dirac delta function. The conditional density of T_{Si} for given $T_W = t_W$ is written as

$$r_{Si}(t \mid t_W) = \sum_{n=0}^{\infty} p_i(n \mid t_W) r_{Si|n}(t) \tag{2.19}$$

For advection-dominated flow, the uptake probability can be approximated as $_x$, which is proportional to the mean time on the surface zone.

Under the assumption of the probability of uptaken into storage is to be unaffected by storage history and mutually independent, the conditional density of T_S given $T_W = t_W$ is, in which $i = 1, 2, ..., N_S$

$$r_S(t \mid t_W) = r_{S1}(t \mid t_W) * ... * r_{SN_S}(t \mid t_W) \tag{2.20}$$

When a particle moving along surface is not modified by entrapment within the storage, the residence time in flow zone is given by

$$r_W(t; x) = \frac{x}{2\sqrt{\pi K_f t^3}} \exp\left[-\frac{(x - Ut)^2}{4K_f t} \right] \tag{2.21}$$

Marion et al. (2008) derived Eq (2.21) from a solution of conventional ADE for a mass injection. Using preceding equations, the overall residence time is expressed as convolution of storage residence time distribution (RTD) and flow zone RTD as following equations

$$r(t;x) = \int_0^t r_W(t-\tau;x) r_S(\tau \mid t-\tau) d\tau \tag{2.22}$$

When it is assumed that total storage time is dependent on the flow distance, $r_S(t \mid t_W)$ can be replaced with $r_S(t \mid x)$. Then Eq (2.22) can be written alternatively as

$$r(t;x) = r_W(t;x) * r_S(t;x) \tag{2.23}$$

For a stream with uniformly distributed transient storage, conditional distribution of n the entrapment at the time t_W into i the storage can be regarded as Poisson distribution

$$p_i(n \mid t_W) = \frac{(\alpha_i t_W)^n}{n!} e^{-\alpha_i t_W} \tag{2.24}$$

Uptake probability can be thought as a function of the travel distance, hence,

$$p_i(n \mid t_W) = \frac{(\alpha_i x/U)^n}{n!} e^{-\alpha_i x/U} \tag{2.25}$$

where, $\alpha_i = q_L / h$ is the relevant rate [1/T]; q_L is average lateral inflow per unit bed area

[L/T]. Finally, overall RTD Eq (2.22) can be arranged through Laplace transform in convenient form (Margolin et al., 2003).

The STIR has been applied in storage of biofilm and streambed interface (Bottacin-Busolin et al., 2009) successfully. However likewise other methods, STIR can yield nonunique solutions in case of a number of domains are considered with no closure solutions are known (Boano et al., 2014).

2.1.2.9 The Aggregated Dead Zone Model (ADZ)

Davis and Atkinson (2000) tried to simplify the 1Z-TSM, for the conceptual case of pure advection without turbulent mixing in main flow zone area, by presenting an Aggregated Dead Zone (ADZ) model. TSM equations, Eq (2.26) and (2.27), were simplified as Eq (2.28) under the assumption of bulk flow at free flow zone, $K_f = 0$.

$$\frac{\partial C_f}{\partial t} = -U\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + -\frac{1}{\chi^2}C_S \tag{2.26}$$

$$\frac{\partial C_s}{\partial t} = \frac{\chi^2}{\tau}(C_f - C_s) \tag{2.27}$$

$$\frac{\partial C_f}{\partial t} + U\frac{\partial C}{\partial x} = -\frac{1}{\tau}(C_f - C_s) \tag{2.28}$$

for the instantaneous injection of mass M, Davis and Atkinson (2000) derived solution of Eq (2.28) as follows.

$$C_f(x,t) = \frac{M}{A_f U}\delta(t - x/U)e^{-t/\tau} + \frac{M}{A_f U}H(t - x/U)e^{-\frac{\chi^2}{\tau}(t-x/U)}e^{-\frac{x}{U\tau}} \cdot \frac{\chi}{\tau}\left(\frac{x/U}{t - x/U}\right)^{1/2} I_i\left[2\frac{\chi}{\tau}\sqrt{\frac{x}{U}(t - x/U)}\right] \quad (2.29)$$

where $I_1(x)$ is a modified Bessel function of first order and first kind; $H(x)$ is the Heaviside step function of $_x$; $\delta(x)$ is the Dirac delta function; $_\tau$ is the characteristic time scale of solute exchange between storage zone and flow zone; and χ is the dead zone parameter which is equivalent to $\chi^2 = A_f / A_s$. More detailed derivation is given by Davis and Atkinson (2000), and Davis (1991).

Table 2.3 Summary of models accounting for transient storage

Model	Governing Equation	Characteristics	# of key parameters
ADE	$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} = K\frac{\partial^2 C}{\partial x^2}$	Asymptotic outcome of continuous motion with a relatively narrow velocity distribution and mixing (Brownian motion yielding Taylor Dispersion, which is mathematically equivalent to Fickian diffusion with drift).	2
1Z-TSM	$\frac{\partial C_f}{\partial t} = -\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f) + \alpha(C_S - C_f) - \lambda_f C_f$ $\frac{dC_s}{dt} = \alpha\frac{A_f}{A_s}(C_f - C_s) - \lambda_s C_s$	Fickian in-stream transport + first-order mass transfer with well-mixed stationary zones (preasymptotic Brownian motion) Most widely used due to simplicity, while uncertain explanation of key process between surface and subsurface	4

Table 2.3 Summary of models accounting for transient storage (continued)

Model	Governing Equation	Characteristics	# of key parameters
2Z-TSM	$\frac{\partial C_f}{\partial t} = -\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f)$ $+ \alpha_{STS}(C_{STS} - C_f) + \alpha_2(C_{HTS} - C_f) - \lambda_f C_f$ $\frac{dC_{STS}}{dt} = \alpha_{STS}\frac{A_f}{A_{STS}}(C_f - C_{STS}) - \lambda_{STS}C_{STS}$ $\frac{dC_{HTS}}{dt} = \alpha_{HTS}\frac{A_f}{A_{HTS}}(C_f - C_{HTS}) - \lambda_{HTS}C_{HTS}$	Fickian in-stream transport + first-order mass transfer with well-mixed stationary zones. Considering STS and HTS separately	4 + 2 *(# of added storage)
CTRW	$\frac{\partial C}{\partial t} + U\frac{\partial C}{\partial x} - K\frac{\partial^2 C}{\partial x^2} = J_s$	Brownian or Lévy in-stream motion represented by jump length distribution + storage times represented by memory function.	Depending on memory function
MADE	$K\frac{\partial^2 C}{\partial x^2} - U\frac{\partial C}{\partial x} = (1 - \eta + k)\frac{\partial C}{\partial t}$	Expressing boundary layer stagnant zone of STS. Single BTC is needed for calibration.	3

Table 2.3 Summary of models accounting for transient storage (continued)

Model	Governing Equation	Characteristics	# of key parameters
FADE	$\frac{\partial(-J)}{\partial x} - U\frac{\partial C}{\partial x} = \frac{\partial C}{\partial t}$ $J = -K_F \frac{\partial^{F-1} C}{\partial x^{F-1}}$	Asymptotic outcome of motion that is effectively discontinuous in space or time, i.e., involving large displacements or long waiting times relative to the scale of averaging (Lévy motion and/or subordinated motion, yielding anomalous diffusion).	3 (S-FADE, T-FADE) or 4 (ST-FADE)
MRMT	$J_s = \frac{\partial C}{\partial t} g(t) = \int_0^t \frac{\partial C(x, t-\tau)}{\partial \tau} g(\tau) d\tau$ $g(t) = \beta_{tot} \int_{\alpha \min}^{\alpha \max} \frac{(k-2)\alpha^{k-2} e^{-\alpha t}}{\alpha_{\max}^{k-2} - \alpha_{\min}^{k-2}} d\alpha$	Fickian in-stream transport + storage times represented by memory function	Depending on memory function
ASP	$\frac{\partial n C_s}{\partial t} + \frac{1}{A}\frac{\partial A V_\eta C_s}{\partial \eta} = 0$ $J_s = \frac{1}{2}\int_0^\infty f(T)\frac{P}{A_f}\xi[-V_\eta(\tau,T)\|_{\tau=0} C_f + (V_\eta(\tau,T)C_s)\|_{\tau=T}]dT$	Fickian in-stream transport + storage times represented by residence time distribution.	Depending on residence time distribution
STIR	$r(t;x) = \int_0^t r_W(t-\tau;x) r_S(\tau \mid t-\tau) d\tau$	Fickian in-stream transport + storage times represented by zone-specific residence time distributions.	Depending on how zones are specified
ADZ	$\frac{\partial C_f}{\partial t} + U\frac{\partial C}{\partial x} = -\frac{1}{\tau}(C_f - C_s)$	Pure advective in-stream transport + first order mass exchange process	2

2.2 Empirical Equations for Predicting Transient Storage Model Parameters

Thackston and Schenelle (1970) proposed storage zone area ratio prediction formula with hydraulic data. Proposed non-linear equation is derived from relation between storage area ratio $\varepsilon = A_s / A_f$ and friction factor f as follows.

$$\varepsilon = l + mf^n \tag{2.30}$$

Where, l, m, and n are coefficient, and friction factor can be calculated with following equation.

$$f = 8\left(\frac{U_*}{U}\right)^2 \tag{2.31}$$

where, friction velocity $U_* = \sqrt{gR_hS}$ [L/T]; R_h is hydraulic radius [L]; and S is energy slope.

Using least square method, Thackston and Schnelle (1970) proposed regression model about the ratio of flow zone area and storage zone area.

$$\varepsilon = 0.0152 + 0.89 f^{2.22} \tag{2.32}$$

Pedersen (1977) proposed formulae for predicting storage zone model parameters K_f, ε, and T. Pedersen (1977) collected hydraulic, geographic, concentration distribution data from the data set of Nordin and Sabol (1974). To predict storage zone parameters, Pedersen (1977) developed a prediction model with dimensionless terms that have physical meaning. He proposed equations for estimating K_f, ε, and T using time-varying concentration distribution data with the determined coefficient C_K.

$$T_L = 0.07\frac{W^2}{hU_*} \tag{2.33}$$

$$K_f = C_K\frac{W^2U_*}{h}\left[1+10^2\, f(1-\frac{f'}{f})\right] = (0.22+35.6f)\frac{W^2U_*}{h} \tag{2.34}$$

$$T = 5f\,\frac{W^2}{hU_*} \tag{2.35}$$

$$\varepsilon = 2f\left(1-\frac{f'}{f}\right) = 2(0.89f-0.0045) \tag{2.36}$$

Pedersen (1977) suggested that the coefficient C_K is in the range of 0.1 ~ 0.4.

Bencala and Walters (1983) estimated the storage zone model parameter matching concentration distribution data with the numerical simulation result. Bencala and Walters (1983) showed storage zone area ratio has a relationship with friction factor using previous tracer test data (Thackston and Schnell,1970; Pedersen, 1977; Nordin and Troutman, 1980).

$$\varepsilon = C_{\varepsilon} f^{-1/2} \tag{2.37}$$

The prediction equation of area ratio was derived by Seo and Yu (2000) derived in terms of friction factor as follows.

$$\varepsilon = 0.737 \exp(-0.945 f^{-0.5}) \tag{2.38}$$

Cheong and Seo (2003) conducted nonlinear regression on published field data using weighted one-step Huber method. Expressions of formulae for TSM parameters were in dimensionless form as follows.

$$\frac{K_f}{hU_*} = 0.583 \left(\frac{U}{U_*} \right)^{0.562} \left(\frac{W}{h} \right)^{1.287} \tag{2.39}$$

$$\frac{T}{h/U_*} = 56.68 \left(\frac{U}{U_*} \right)^{-0.884} \left(\frac{W}{h} \right)^{0.767} \tag{2.40}$$

$$\varepsilon = 0.85 \left(\frac{U}{U_*} \right)^{-0.401} \left(\frac{W}{h} \right)^{0.008} \tag{2.41}$$

Cheong et al. (2007) collected published data (Fischer et al., 1968; Godfrey and Frederick, 1970; Nordin and Sabol, 1974; Graf, 1995; Czernuszenko et al., 1998) for estimate key

parameters of TSM. To predict storage exchange in natural streams, Cheong et al. (2007) analyzed data and developed equations in non-dimensional terms. The importance of Peclet number and channel sinuosity was pointed out in transient storage solute exchange mechanisms. The nonlinear multiple equations were proposed using robust minimum covariance determinant method as follows.

$$\frac{K_f}{hU_*} = 43.928\left(\frac{U}{U_*}\right)^{0.8453}\left(\frac{W}{h}\right)^{0.2176}\left(\frac{UL}{K}\right)^{-0.0420} S_n^{-1.6891} \tag{2.42}$$

$$\frac{T}{h/U_*} = 20.595\left(\frac{U}{U_*}\right)^{-1.4625}\left(\frac{W}{h}\right)^{0.6639}\left(\frac{UL}{K}\right)^{0.3232} S_n^{1.9132} \tag{2.43}$$

$$\varepsilon = 0.8279\left(\frac{U}{U_*}\right)^{-0.3550}\left(\frac{W}{h}\right)^{-0.2590}\left(\frac{UL}{K}\right)^{-0.2759} S_n^{0.3315} \tag{2.44}$$

where, S_n is the channel sinuosity.

Sahay (2012) developed empirical equations for TSM parameters using published 58 datasets of 1D tracer test in 33 natural streams, summarized by Cheong et al. (2007). In the study of Sahay (2012), TSM prediction equations in dimensionless form were derived using a genetic algorithm (GA) optimization technique.

$$\frac{K_f}{hU_*} = 0.41\left(\frac{U}{U_*}\right)^{0.98}\left(\frac{W}{h}\right)^{0.663}\left(\frac{UL}{K}\right)^{-0.777} S_n^{2.226} \tag{2.45}$$

$$\frac{T}{h/U_*} = 15.62\left(\frac{U}{U_*}\right)^{-1.897}\left(\frac{W}{h}\right)^{1.698}\left(\frac{UL}{K}\right)^{0.797} S_n^{0.986} \tag{2.46}$$

$$\varepsilon = 0.04853\left(\frac{U}{U_*}\right)^{-0.1087}\left(\frac{W}{h}\right)^{-0.0048}\left(\frac{UL}{K}\right)^{-0.1459} S_n^{-0.0749} \quad (2.47)$$

Recently, empirical equations for 3 of OTIS based TSM parameters (K_f, A_s, and α) were proposed by Femeena et al. (2019). They carried out meta-analysis to construct datasets for the derivation of TSM empirical equations. The maximum number of points, for calibration and validation, was 380. The equations were developed under the consideration of the uncertainty of measurement. Even if the researchers had tried to overcome a problem that an estimated parameter set is feasible only for a certain tracer test data, there are limitations that developed equations do not have generalized form and dimensions of both side are not identical. The simple equations were derived manually, trial and error method, as follows.

$$K_f = 1.5UWh^{0.5} \quad (2.48)$$

$$\alpha = \frac{0.001U}{Wh} \quad (2.49)$$

$$A_s = 0.1\left[0.1W + \frac{Q}{h}\right]^{1.2} \quad (2.50)$$

Table 2.4 Summary of empirical equations predicting transient storage parameters

Reference	Equation	Derivation method
Thackston and Schenelle (1970)	$\varepsilon = 0.0152 + 0.89 f^{2.22}$ $f = 8\left(\frac{U_*}{U}\right)^2$	Data fitting
Pedersen (1977)	$K_f = C_K \frac{W^2 U_*}{h}\left[1 + 10^2 f(1 - \frac{f'}{f})\right] = (0.22 + 35.6 f)\frac{W^2 U_*}{h}$ $T = 5 f \frac{W^2}{h U_*}$ $\varepsilon = 2 f \left(1 - \frac{f'}{f}\right) = 2(0.89 f - 0.0045)$	Data fitting
Bencala and Walters (1983)	$\varepsilon = C_\varepsilon f^{-1/2}$	Data fitting
Seo and Yu (2000)	$\varepsilon = 0.737 \exp(-0.945 f^{-0.5})$	Data fitting

Table 2.4 Summary of empirical equations predicting transient storage parameters (continued)

Reference	Equation	Derivation method
Cheong and Seo (2003)	$\frac{K_f}{hU_*} = 0.583\left(\frac{U}{U_*}\right)^{0.562}\left(\frac{W}{h}\right)^{1.287}$ $\frac{T}{h/U_*} = 56.68\left(\frac{U}{U_*}\right)^{-0.884}\left(\frac{W}{h}\right)^{0.767}$ $\varepsilon = 0.85\left(\frac{U}{U_*}\right)^{-0.401}\left(\frac{W}{h}\right)^{0.008}$	Weighted one-step Huber method for nonlinear regression
Cheong et al. (2007)	$\frac{K_f}{hU_*} = 43.928\left(\frac{U}{U_*}\right)^{0.8453}\left(\frac{W}{h}\right)^{0.2176}\left(\frac{UL}{K}\right)^{-0.0420} S_n^{-1.6891}$ $\frac{T}{h/U_*} = 20.595\left(\frac{U}{U_*}\right)^{-1.4625}\left(\frac{W}{h}\right)^{0.6639}\left(\frac{UL}{K}\right)^{0.3232} S_n^{1.9132}$ $\varepsilon = 0.8279\left(\frac{U}{U_*}\right)^{-0.3550}\left(\frac{W}{h}\right)^{-0.2590}\left(\frac{UL}{K}\right)^{-0.2759} S_n^{0.3315}$	Weighted Robust Minimum Covariance Determinant Method
Sahay (2012)	$\frac{K_f}{hU_*} = 0.41\left(\frac{U}{U_*}\right)^{0.98}\left(\frac{W}{h}\right)^{0.663}\left(\frac{UL}{K}\right)^{-0.777} S_n^{2.226}$ $\frac{T}{h/U_*} = 15.62\left(\frac{U}{U_*}\right)^{-1.897}\left(\frac{W}{h}\right)^{1.698}\left(\frac{UL}{K}\right)^{0.797} S_n^{0.986}$	Genetic Algorithm

$$\varepsilon = 0.04853\left(\frac{U}{U_*}\right)^{-0.1087}\left(\frac{W}{h}\right)^{-0.0048}\left(\frac{UL}{K}\right)^{-0.1459} S_n{}^{-0.0749}$$

Table 2.4 Summary of empirical equations predicting transient storage parameters (continued)

Reference	Equation	Derivation method
Femeena et al. (2019)	$K_f = 1.5UWh^{0.5}$ $\alpha = \frac{0.001U}{Wh}$ $A_s = 0.1\left[0.1W + \frac{Q}{h}\right]^{1.2}$	Data fitting

2.3 Parameter Estimation

Water resource problems have mathematical features: nonlinearity, stochasticity, discreteness, non-convexity, high dimension decisions, severe combinatorial growth rates, and uncertainty (Reed et al. 2013). Due to pointed properties, researchers have considered it is effective that use multi-objective evolutionary algorithms.

Many studies of multi-objective optimization performance were conducted for managing water resources (Haimes and Hall, 1977; Cohon and Marks, 1975; Reed et al., 2013). Reed et al. (2013) assessed about performances of modern evolutionary multi-objective optimization techniques. Among the analyzed techniques, adaptive operator selection method performed consistently well at competitive to superior performance levels on all of the analyzed problems on the study.

Previous works were conducted to deal with estimating TSM parameter using multi-objective optimization algorithm, to identify feature of river stream. The TSM parameters were evaluated with multi-layer perceptron artificial neural networks using dataset obtained by Cheong et al. (2007) (Rowiński and Piotrowski, 2008). In addition, three different optimization techniques for training neural networks, namely the Levenberg-Marquardt algorithm, particle swarm optimization, and differential evolution methods, were compared on applicability. The differential evolution method showed better performance than Levenberg-Marquardt algorithm and particle swarm optimization. However, differential evolution is considerably more time consuming (Rowiński et al., 2008; Naeini et al., 2018). Kerr et al. (2012) merged Matlab coded OTIS with SCE-UA algorithm to investigate effect of 2 storage zone TSM mass exchange structure. A software assessing uncertainty in TSM parameter using Monte Carlo simulation was proposed by Ward et al. (2017). The software, which is named OTIS-MCAT, used OTIS

for solute transport simulation.

Naeini et al. (2018) proposed global optimization framework named as Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL). The SC-SAHEL is a parallel optimization framework which is based on the SCE-UA algorithm architecture. The SC-SAHEL partition population into different complexes as equally sized groups and update with each evolutionary algorithm before shuffle complexes as presented flowchart (Figure 2.4). This framework provides adaptive multiple evolutionary algorithms at evolution step. During adaptation process, SC-SAHEL assigns number of complex to each evolutionary method in which the algorithm with best performances will have additional complex to evolve in the next iteration, while the other worst algorithm loses complex. In this framework, four evolutionary algorithms were accorded, modified competitive complex evolution, modified frog leaping, modified grey wolf optimizer, and differential evolution.

Even though previous works showed adaptive algorithm guarantees consistent high performance on handling water resource problem, applicability of EA in certain problem is different by case (Wolpert and Macready, 1997). Therefore compatibility of calibration method was assessed in Chapter 3 by changing EA of SC-SAHEL. The SC-SAHEL is adaptive optimization framework that exploits multiple evolutionary algorithms, including differential evolution algorithm. SC-SAHEL select combination of provided EAs. Thus, parameter estimation using various EA combinations can be performed using SC-SAHEL such as SCE-UA, SP-UCI, or both algorithms. Moreover, this algotirhm provides parallel computing user can find solution much efficiently. As a result, the SC-SAHEL was adopted as main optimization starategy in this study in order to identify compatibility of optimization condition, efficiently.

In Section 2.3 and sub-sections, overview of SC-SAHEL is given.

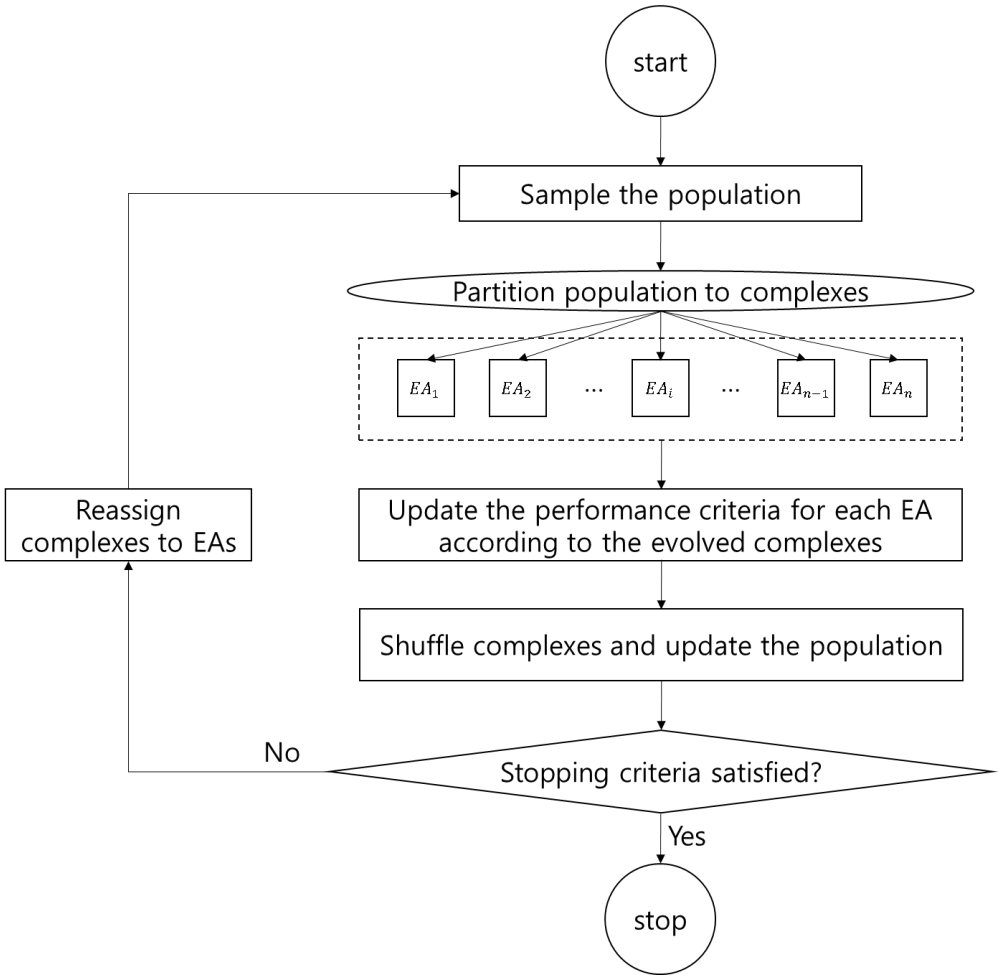


Figure 2.2 The SC-SAHEL framework flowchart (adapted from Naeini et al. 2018)

2.3.1. The SC-SAHEL Framework

Likewise other multi-objective optimization algorithms, e.g. Shuffled complex evolution referred to as SCE-UA, SC-SAHEL optimization framework starts with generating population in feasible space. And the population is subdivided into defined number of complexes. These complexes help to maintain diversity of sampled population. Contrary to SCE-UA that adopts single evolution algorithm, SC-SAHEL assigns different number of complex to each evolutionary method based on calculation of following Evolutionary Method Performance metric.

$$\text{EMP}=\frac{\bar{F}-\bar{F}_N}{\bar{F}} \tag{2.51}$$

where, $\bar{F}$ and $\bar{F}_N$ are mean objective function values before and after evolution in each complex, respectively. SC-SAHEL increases complexes of best performance EAs, and decreases complexes of worst EA depend on EMP metric.

Figure 2.2 is briefly showing flowchart of the SC-SAHEL. The explanation of SC-SAHEL framework flow of which Figure 2.2 is as follows.

Step 1 Sets number of complexes (NGS) proportional to the number of evolutionary algorithms (EAs).

Step 2 Samples population in the feasible parameter space.

Step 3 Calculates objective function value of each parameter point.

Step 4 Ranks and sorts all parameters in order of increasing function values.

Step 5 Separates all values into complexes and assigns complexes to EAs.

Step 6 If a user wants to consider population dimensionality, it can be restored using the PCA algorithm.

Step 7 Evolves each complex with matched EAs.

Step 8 Calculates mean EMP for each EA

Step 9 Reassigns the number of the complex by the calculated value of EMP.

Step 10 Shuffle complexes.

Step 11 If the convergence criteria are not satisfied, go to Step 3, otherwise, stop.

Five different EAs are employed into the SC-SAHEL for adaptive assignment of population. A user can choose EAs for optimization among adopted various EAs, Competitive Complex Evolution (CCE) of SCE-UA (Duan et al., 1993), Modified Competitive Complex Evolution (MCCE) of SP-UCI (Chu et al., 2011), Modified Frog Leaping (MFL) which is Particle Swarm Optimization (PSO) based algorithm (Eusuff and Lansey, 2003), Modified Grey Wolf Optimizer (GWO) proposed by Mirjalili et al. (2014), Modified Differential Evolution (DE) integrated into SCE-UA (Mariani et al., 2011).

Exploited EAs were simply explained trough following sub-sections. Further details of each algorithm, such as pseudo code and modification information, can be found in the work of Naeini et al. (2019)

2.3.1.1 Modified Competitive Complex Evolution (MCCE)

The CCE and MCCE are Nelder-Mead Simplex method (Nelder and Mead, 1965) based EAs. The CCE is regarded as robust, efficient, and effective EA. However Chu et al. (2011) found that "population degeneration" phenomenon. It is referred to when the SCE-UA does not converge on high dimensional problem losing parameter span. This phenomenon causes shrinkage into local optimum. By employing Principle Component Analysis (PCA), Chu et al. (2011) modified CCE algorithm in order to restore the parameter space dimension. By adopting MCCE into Shuffled complex strategy, Chu et al. (2011) developed SP-UCI.

2.3.1.2 Modified Frog Leaping (MFL)

The Frog Leaping (FL) algorithm is PSO based local search mechanism adopted within SCE-UA framework by Eusuff and Lansey (2003). Eusuff et al. (2006) showed FL is efficient for the discrete optimization problem and converges to optimal value much faster than GA algorithm. Naeini et al. (2018) modified FL in order to apply FL to continuous domain and adopt SC-SAHEL parallel framework. Basically FL uses best point of complex and entire population, however, MFL uses best point of subcomplex to avoid interaction with other EAs. Third difference is MFL uses two different jump rate for better exploration ability during optimization process. If performance of offspring is not better than performance of parents, ML randomly selects value within whole population, instead, MFL finds within the range of subcomplex for search speed.

2.3.1.3 Modified Grey Wolf Optimizer (GWO)

Inspired by social behavior of grey wolves, the way of hunting and hierarchy, Mirjalili et al. (2014) developed meta-heuristic algorithm named as The Grey Wolf Optimizer. When grey wolves hunt, at first they chase and approach to the prey. And wolves encircle and pursue the prey. Finally, wolves attack the prey. Likewise hunting steps of grey wolves GWO find three top fitted points. The points contribute to evolution so that GWO navigates population to best solution. Mirjalili et al. (2014) showed GWO is efficient and effective in many problems compared to popular algorithms like PSO and GA. While GWO uses adaptive jump algorithm, SC-SAHEL adopted modified GWO (MGWO) using three jump rate in order to limit information to assigned complex in SC-SAHEL.

2.3.1.4 Modified Differential Evolution (DE)

The Differential Evolution was proposed by Storn and Price (1997). DE is regarded as powerful but simple heuristic algorithm (Qin and Suganthan, 2005; Sadegh and Vrugt, 2014; Naeini et al., 2018). The DE was integrated into SCE-UA by Mariani et al. (2011). It was shown that DE can give robust solutions compared to SCE-UA. However DE has slow performance in comparison to other EAs, so Naeini et al. (2018) modified DE adding three different mutation rate. At first, the algorithm uses large mutation rate. Then uses quarter of first mutation rate and goes to half of first step. If no better offspring were generated, the new points are selected randomly in range of complex.

2.4 Regression Methods

As described in Table 2.4, empirical equations for TSM parameters were developed in the least square data fitting. Cheong and Seo (2003) used nonlinear multi-variable weighted one-step Huber regression. The weighted robust minimum covariance determinant was adopted in Cheong et al. (2007)'s equations. Both of the methods are one of the least square outlier detection methods. Especially, the weighted robust minimum covariance determinant method ignores outlier by giving zero weight to the observations. Since same large weight of outlier can cause over-fitting. These outlier detection techniques have the advantage of overcoming the over-fitting problem. Recently, data-driven regression techniques are widely used. Sahay (2012) developed equations with the GA optimization technique. In order to avoid over-fitting outliers were ignored while deriving equations. In the same manner, outlier detection will be considered in this study.

For the prediction of one-dimensional longitudinal dispersion coefficient, recently, many of data-driven methods are being used. Etemad-Shahidi and Taghipour (2012) used the M5 model tree method. The M5 algorithm is one of the most commonly used model tree method. Sattar and Gharabaghi (2015) used Gene expression. Alizadeh et al. (2017) used the Particle Swarm Optimization (PSO) algorithm. Artificial Neural Networks (ANN) models have been also adopted in the prediction of one-dimensional dispersion coefficient (Rowiński et al.2005; Tayfur and Singh 2005; Toprak and Cigizoglu 2008; Sahay 2011). These methods are powerful in finding constants and it tends to produce functions that grow in length over time (Davidson et al.1999, 2000).

Mostly, data-driven techniques are called a black box model. These black box models have ability to deal with non-linear process came from model complexity. However, the models

require some tuning of parameter of itself. Symbolic regression methods have advantages over other machine learning methods such as neural networks and Support Vector Machines (SVMs) in case of that user need interpretable or generalized analytic solutions (Smits et al., 2005; Vladislavleva et al., 2009). Giustolisi and Laucelli (2005) pointed out that calibration of parameters, over-fitting problems and lack of knowledge about the learning process are a disadvantage of black-box data-driven techniques. In order to overcome disadvantage of black-box models and conventional data fitting methods, previous efforts tried to develop symbolic regression methods represented as Grey-box models (Koza, 1992; Giustolisi and Savic, 2006; Searson, 2010, 2015). As described in Chapter 1, Multi-Gene Genetic Programming (MGGP) is widely used and has been giving satisfactory result in civil engineering problem. Furthermore, it does not require a specific form of formula contrary to other methods.

1SZ-TSM describes not only one transient storage, but also a number of HTS and STS, which are governed by each different solute exchange process, via one storage exchange equation. This lumped model characteristic implies storage zone parameters of 1SZ-TSM only could be expressed with non-linear summations of input hydromorphic variables. For that reason, Femeena et al. (2019) presented simple non-linear equations. Taking non-linearity of the process into account, it is expected that MGGP based formulae can express complexities of natural streams, such as transient storages. Symbolic regression techniques can be a substitution of classical statistical methods in this problem as portrayed by Vladislavleva et al. (2009). Accordingly, the MGGP was applied to derive prediction formulae of TSM parameters.

In the following sub-sections, the introduction of MGGP and a similar method of MGGP, Evolutionary Polynomial Regression (EPR), are presented.

2.4.1. The Multi-Gene Genetic Programming (MGGP)

2.4.1.1 The Simple Genetic Programming

Genetic Programming (GP) is a specialized Genetic Algorithm (GA), which is an evolutionary technique mimicking natural evolution processes such as mutation and crossover. Koza (1992) firstly put into use GP as a symbolic regression technique in many engineering problems. The main difference of GP to conventional GA is that thee GP evolves tree-based data structures while the GA evolves number strings (Gandomi and Alavi, 2012). The GP attempts to find the fittest optimal model, by constructing and modifying trees consist of functions and variables.

Standard GP creates a random population of genes. Created genes are consist of functions and terminals as a tree structure. Function parts, root and linked branches of the tree, can be any mathematical (+,−,×, ÷, tanh, etc.) or Boolean logic functions (If-then-else, OR, Greater than, etc.). The leaves of tree, terminals, can assign numerical constants, input variables of the target model, etc. (Gandomi et al., 2010). In this manner, the mathematical form of the standard GP model can be expressed by combining these functions and terminals of tree data structure.

Algorithm of GP adopted Darwin's natural selection theory to the evolution of the population, likewise in an evolutionary feature of GA. At first, a random population of an individual gene is generated. Once the population of genes is randomly generated and fitness function values are evaluated, each gene is modified by the principle of natural evolution with mutation and crossover, reproducing offspring. For the crossover operation, terminals or branched nodes of parent trees are randomly selected, and the selected points are exchanged, as depicted in Figure 2.3. Additionally, the mutation process picks branches, along with sub-nodes, and replace each

bunch with randomly generated subtree (Figure 2.4). The crossover and the mutation operations are applied, to models with low fitness, after fitness values of each function are evaluated. Koza (1994) suggested that the ratio of reproduction, crossover, and mutation is 85:10:5, respectively. This evolution step is iterated until the termination criterion is met enhancing model fitness of the produced models from the GP.

2.4.1.2 Scaled Symbolic Regression via Multi-Gene Genetic Programming

The MGGP is a scaled symbolic regression method that is an advanced version of the standard GP (Searson, 2010). The MGGP model is a linear weighted combination model of gene trees produced from the standard GP. Contrary to the GP produces regression models with each gene tree, the MGGP uses one or more of gene trees and calibrate coefficients of gene trees using statistical regression methods such as least-square regression. Typical MGGP model can be illustrated as Figure 2.5 for instance, where the b_i is the coefficients of each gene tree.

In the same way as classic GP that iteratively reproduces new models by the crossover and mutation, the the algorithm of the MGGP contains the crossover and mutation as well. The crossover process of MGGP is a so-called high-level crossover (in standard GP, it is called as a low-level crossover). In the MGGP, individual gene trees, which are from the GP's generation process, are regarded as a gene. Thus, the high-level crossover is performed on the tree level (Searson, 2010). Taking the example for a two-point high-level crossover, if two parent models, $(T_1T_2[T_3T_4])$ and $(T_5[T_6]T_7)$, are exist in a certain step, components in '[ ]' brackets are selected randomly, then reproduced children become $(T_1T_2[T_6])$ and $(T_5[T_3T_4]T_7)$. Similar to the high-level crossover, the high-level mutation process mutates randomly selected trees.

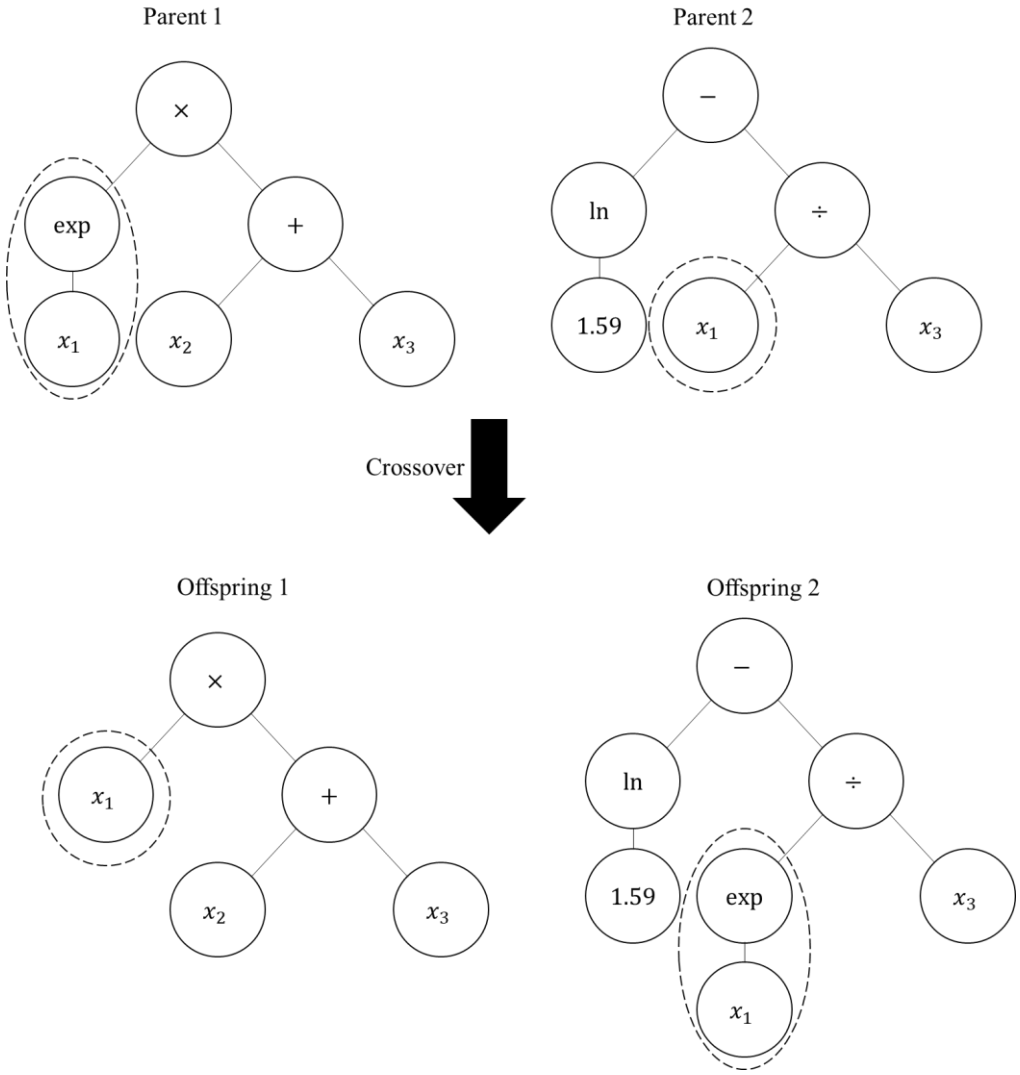


Figure 2.3 The example of crossover operation in GP

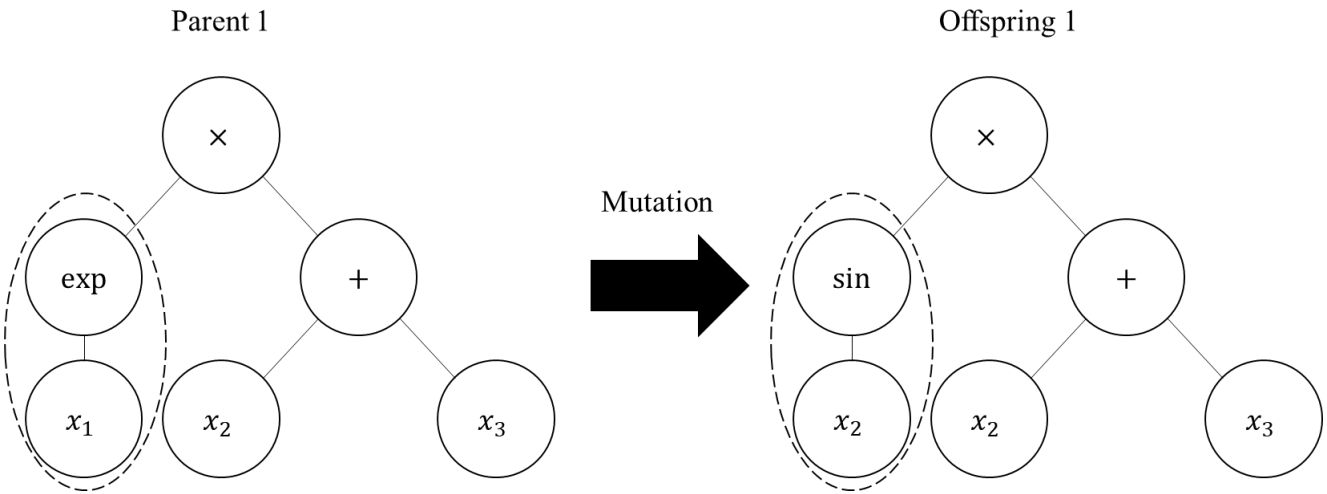


Figure 2.4 The example of mutation operation in GP

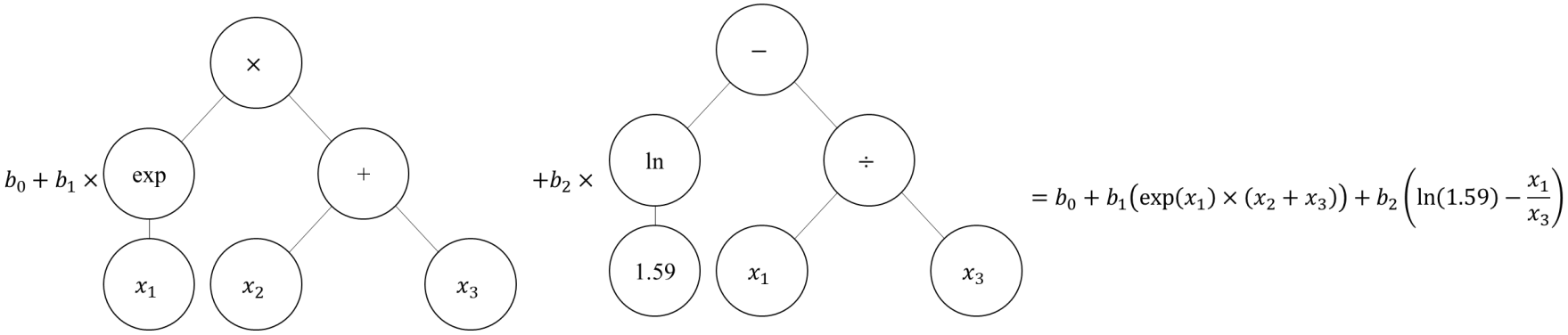


Figure 2.5 The example of MGGP model

2.4.2. Evolutionary Polynomial Regression (EPR)

The Evolutionary Polynomial Regression (EPR) is a hybrid data-driven technique grey-box model developed by Giustolisi and Savic (2006). It is introduced as a nonlinear stepwise regression method. The idea of EPR came from the new regression method of Davison et al. (1999), so-called Rule-based symbolic regression (R-BSR). They incorporated Genetic Programming (GP) and least squares optimization, to find a form of polynomial expressions and constant value of formula, respectively. Based on the idea of the R-BSR, Giustolisi and Savic (2006) combined GA and LS. The primitive model of EPR used Single objective genetic algorithm (SOGA) (Goldberg, 1989). Due to some drawbacks, performance decreasing effect as increase number of terms and discard of less term with high performance, Giustolisi and Savic (2009) adopted a Multi-objective genetic algorithm (MOGA) (Goldberg, 1989; Giustolisi et al., 2004). The objectives of adopted MOGA are (1) maximizing model accuracy, (2) minimizing the number of polynomial coefficients, and (3) minimizing the number of the input variable. The EPR is able to generate polynomial models of multiplication of functions where each term consist of candidate variables. Each variable has own exponent constructed from MOGA sequence, and their constant coefficients are estimated with LS. The forms of pseudo polynomial formula can be managed by EPR are Eq. (2.52)-(2.55).

$$\hat{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)} \cdot f\left((X_1)^{\mathrm{ES}(j,1)}\right) \cdot \ldots \cdot f\left((X_k)^{\mathrm{ES}(j,k)}\right) \quad (2.52)$$

$$\hat{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot f\left((X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)}\right) \quad (2.53)$$

$$\hat{Y} = a_0 + \sum_{j=1}^{m} a_j \cdot (X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)} \cdot f\left((X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)}\right) \tag{2.54}$$

$$\hat{Y} = g\left(a_0 + \sum_{j=1}^{m} a_j \cdot (X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)}\right) \tag{2.55}$$

where, $\hat{Y}$ is predicted value of model; m is the number of additive terms; X_j are the candidate inputs; a_j are the constant coefficient of each term; k is the number of input variables; ES is the exponent matrix, the values of ES are accorded with user defined exponent vector EX.

2.4.2.1 Main Flow of EPR Procedure

The EPR procedure is composed of two-stage as follows: (1) searching best function structure using GA, by combining vectors iteratively, and (2) finding constant coefficient values of each terms performing LS regression. The general form of expression in EPR is written as

$$\hat{y} = a_0 + \sum_{j=1}^{m} F(X, f(X), a_j) \tag{2.56}$$

The Eq. (2.56) can be transferred as vector form of

$$\hat{Y}_{N\times 1}(\theta, Z) = \left[I_{N\times 1} \quad Z^{j}_{N\times m} \right] \times \left[a_0 \quad a_1 \quad \ldots \quad a_m \right]^T = Z_{N\times(m+1)} \times \theta^T_{(m+1)\times 1} \tag{2.57}$$

in which, $Z^{j}_{N\times m}$ is product of input variable matrix X :

$$Z^{j}_{N\times m} = \left[(X_1)^{\mathrm{ES}(j,1)} \cdot \ldots \cdot (X_k)^{\mathrm{ES}(j,k)} \right] \tag{2.58}$$

$$X = \begin{bmatrix} x_{11} & x_{12} & \ldots & x_{1k} \\ x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ x_{N1} & x_{N2} & \ldots & x_{Nk} \end{bmatrix} = \left[X_1 \quad X_2 \quad \ldots \quad X_k \right] \tag{2.59}$$

The ES is exponent vector as aforementioned in 2.3. If input parameter is chosen to be $\mathrm{EX} = [-2, 1, 0]$ with the number of terms in equation $m = 3$ and the number of independent inputs $k = 4$, then exponent vector becomes

$$\mathrm{ES} = \begin{bmatrix} -1 & 1 & 2 & 0 \\ 0 & 0 & 1 & -2 \\ 2 & 0 & -1 & 1 \end{bmatrix} \tag{2.60}$$

By putting value of the exponent matrix defined as Eq. (2.59) into the Eq. (2.57), the expression of input product is

$$
\begin{aligned}
Z_1 &= (X_1)^{-1}(X_2)^1(X_3)^2(X_4)^0 = (X_1)^{-1}(X_2)^1 \\
Z_2 &= (X_1)^0(X_2)^0(X_3)^1(X_4)^{-2} = (X_3)^1(X_4)^{-2} \\
Z_3 &= (X_1)^2(X_2)^0(X_3)^{-1}(X_4)^1 = (X_1)^2(X_3)^{-1}(X_4)^1
\end{aligned} \tag{2.61}
$$

Implementing Eq. (2.61), expression of formula is given as following equation

$$
\begin{aligned}
\hat{Y} &= a_0 + a_1 Z_1 + a_2 Z_2 + a_3 Z_3 \\
&= a_0 + a_1 (X_1)^{-1}(X_2)^1 + a_2 (X_3)^1 (X_4)^{-2} + a_3 (X_1)^2 (X_3)^{-1}(X_4)^1 \\
&= a_0 + a_1 \frac{(X_2)^1}{(X_1)^1} + a_2 \frac{(X_3)^1}{(X_4)^2} + a_3 \frac{(X_1)^2 (X_4)^1}{(X_3)^1}
\end{aligned} \tag{2.62}
$$

The constant coefficients of each term a_j are estimated with LS method, which minimizes SSE. Traditionally, this inverse problem is solved by Gaussian elimination. However, the vector $\left[I_{N\times 1} \quad Z^j_{N\times m} \right]$ may not be full rank or columns are linearly dependent. Therefore, Singular value decomposition (SVD) was adopted for the solver of LS problem (Giustolisi and Savic, 2006).

Finding best form of equation such as present Eq. (2.62), the MOGA is implemented in the EPR. The GA is and algorithm inspired by Darwinian evolution. The parameters of GA are optimized using evolution mechanism of 'chromosomes', which is found in DNA. In the EPR, integer GA coding was used to determine the exponent matrix. In the Eq. (2.60) the location of candidate exponents matrix $\mathrm{EX}=\left[-2,-1,0,1,2\right]$ is set as $\left[2\ 4\ 5\ 0,\ 3\ 3\ 4\ 1,\ 5\ 3\ 2\ 4\right]$ using integer GA coding when standard GA expresses binary alphabet. Giustolisi and Savic (2009) noted that, in the implemented GA, following parameters were adopted: "multiple-point

crossover, single-point mutation, the number of maximum polynomial terms j, and the number of inputs k".

Chapter 3. Model Development

3.1 Numerical Model

Previous researches about empirical equations for transient storage model parameters (Cheong et al., 2007; Sahay, 2012) estimated free flow zone velocity U_f using area ratio with the manner of the following equation.

$$U_f = U\frac{(A_f + A_s)}{A_f} = U(1+\varepsilon) \tag{3.1}$$

In some cases, the free flow zone velocity is regarded as a known value. For instance, Femeena et al. (2019) took measured section velocity to the U_f for calculating free flow zone area A_f.

Transient storage model parameters have the effect of both surface transient storage and hyporheic transient storage. Even if two cases of flow zone areas were same, also estimated free flow zone velocities, estimated storage zone area could be different because of non-surface transient storage effect. In other words, area ratio can be different in the same free flow velocity cases. Ward et al. (2018) conducted experimental research in geomorphically similar sites having different restoration status (i.e., restored or unrestored) (Figure 1.1). According to the study, usually in restored reach with gravel have a larger exchange coefficient α, area ratio A_s / A_f, and smaller residence time in the stream T_{str} despite similar geomorphic features. The investigators interpreted that dominant exchange processes and dynamics with discharge

accounting for the differences between the surface transient storage and hyporheic transient storage. Accordingly, the estimation method of the free flow zone velocity with the storage ratio have the possibility of neglecting hyporheic transient storage effect in natural stream. For this reason, it can be useful to estimate the free flow zone area A_f and the storage zone area A_s respectively illustrating hyporheic storage effect. As discussed above and in section 2.1, it is anticipated that 1Z-TSM is an appropriate model for solute transport simulation. Accordingly, 1Z-TSM was employed at the parameter estimation model of this study.

Using TSM equation, Runkel and Chapra (1993) presented numerical model One-dimensional Transport with Inflow and Storage (OTIS). Later, Runkel (1998) employed a non-linear regression technique for parameter estimation to OTIS (OTIS-P). Many researches about transient storage are being conducted using OTIS-P or OTIS based parameter searching technique (Choi et al., 2000; O'conner et al., 2010; Kelleher et al. 2013; Rana et al., 2017; Kerr et al., 2012; Ward et al., 2017; Bohrman and Strauss, 2018; TT Le et al., 2018).

For exhaustive optimization technique, SC-SAHEL, classical TSM was employed for concentration simulation. Iterative simulations were conducted without biochemical reaction to estimate only unknown TSM parameters. The adopted numerical scheme is the Crank-Nicolson method as same as OTIS.

$$\frac{\partial C_f}{\partial t} = -\frac{Q}{A_f}\frac{\partial C_f}{\partial x} + \frac{1}{A_f}\frac{\partial}{\partial x}(A_f K_f \frac{\partial C_f}{\partial x}) + \frac{q_L}{A_f}(C_L - C_f) + \alpha(C_S - C_f) \qquad (3.2)$$

$$\frac{dC_s}{dt} = \alpha \frac{A_f}{A_s}(C_f - C_s) \qquad (3.3)$$

The Finite Volume Method (FVM) for solving TSM equation without decay terms is described by Runkel and Chapra (1993). Boundary and initial conditions are given by

$$K_f \frac{\partial C_f}{\partial x} = 0 \quad \text{at} \quad x = L \tag{3.4}$$

$$C = C_{bc} \quad \text{at} \quad x = 0, t > 0 \tag{3.5}$$

$$C = C_{init} \quad \text{at} \quad t = 0 \tag{3.6}$$

$$C_S = C_{Sinit} \quad \text{at} \quad t = 0 \tag{3.7}$$

Where C_{bc} is the solute concentration at the upper boundary $[\mathrm{M/L^3}]$; C_{init} and C_{Sinit} are initial background concentration of in-stream and storage area, respectively $[\mathrm{M/L^3}]$.

3.1.1. Model Validation

In order to validate the result of the TSM numerical model in this study, two simulation criterions were compared. At first, continuous injection of an 1-D analytical solution (Eq. 3.8) and continuous boundary condition of the numerical model for validation of the advection-dispersion mechanism. Next, since the correct simulation of transient storage exchange mechanism is the most important factor, the simulation results of OTIS (Runkel, 1998) and the TSM model were compared, under the same input variables.

$$C(x,t) = \frac{C_0}{2}\left[erfc\left(\frac{x-Ut}{\sqrt{4Kt}}\right) + erfc\left(\frac{x+Ut}{\sqrt{4Kt}}\right)\exp(Ux/K)\right] \tag{3.8}$$

To assess the accuracy of the numerical model, percent error $\mathrm{RMSE/max(C)}$ was selected as a criterion. Simulation conditions are $C_0 = 10\ \mathrm{g/m^3}$, $dx = 1\ \mathrm{m}$ $dt = 1\ \mathrm{sec}$, $U = 1\ \mathrm{m/s}$, and $K = 1\ \mathrm{m/s^2}$. Under the aforementioned numerical condition, analytical solution and replica of OTIS were simulated as Figure 3.1. According to the simulation result, numerical truncation error, calculated as $\mathrm{RMSE/max(C)}$, was 0.0013% of total mass along injection time.

In the comparison of the OTIS and coded model in this study, percent error of simulation result, $\mathrm{RMSE/max(C)}$, was OTIS errors were 0.0086% and 0.006% at 60 m and 98 m, respectively (Figure 3.2).

From two accuracy assessment, the coded model in this study is validated. In other words,

TSM parameter estimation based on the model is identical to the estimated value of OTIS. Accordingly, a meta-analysis was conducted using both published researches and parameter estimation values using the numerical model in this study.

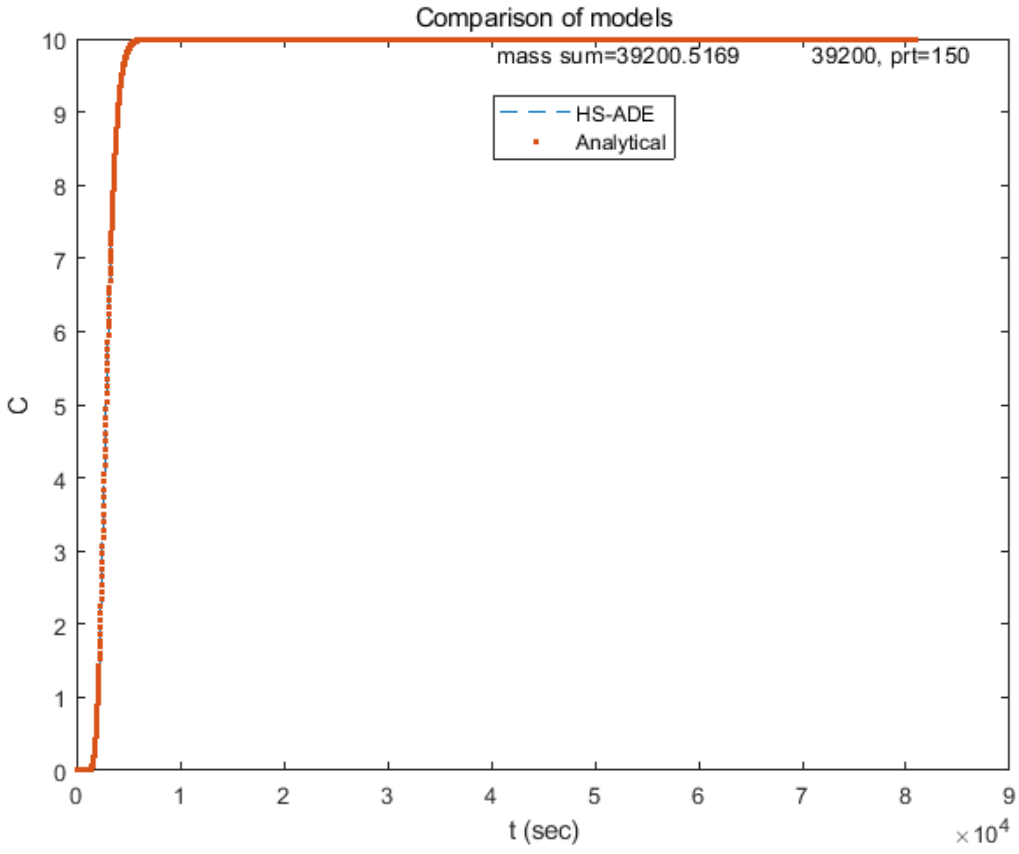


Figure 3.1 Comparison of the 1D-ADE analytical solution and the TSM model (error = 0.0013%)

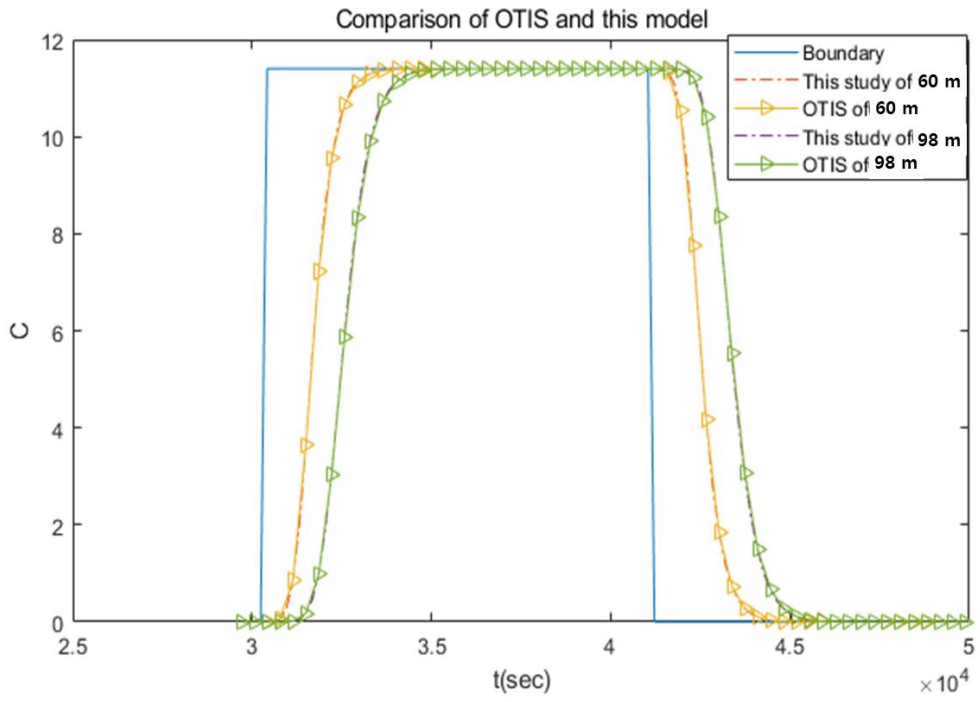


Figure 3.2 Plot of comparison of the TSM model and the OTIS.

(Data from Bencala and Walters (1983))

3.2 Merger of TSM-SC-SAHEL

In this study, 1Z-TSM was integrated with an SC-SAHEL optimization technique.

Measured concentration curve and geomorphic data are necessary to estimate the TSM parameter.

The flowchart of TSM-SC-SAHEL is briefly shown in Figure 3.3. The framework is designed as the following steps:

Step 1 Perform mesh independency test of TSM to obtain optimal value of mesh size yielding reliable numerical solution.

Step 2 Determine optimization variables, i.e. population size, stopping criteria, EAs and so on.

Step 3 Optimize the TSM parameters using SC-SAHEL with objective function of functionalized TSM numerical solution.

Optimization problem of TSM-SC-SAHEL can be formulated as follows

Minimize MSE (3.9)

subject to $0 \leq K_f \leq 3\,0K$ (3.10)

$$0 \leq A_f \leq 4Wh \quad (3.11)$$

$$0 \le A_s \le 4Wh \tag{3.12}$$

$$0 \le \alpha \le 10 \tag{3.13}$$

In the constraint of flow zone dispersion coefficient, empirical equations of one-dimensional dispersion coefficient are used. Because the empirical equation has uncertainty, value upper boundary of K_f was conservatively determined. Likewise, other constraints of the optimization problem were set in wide range of arbitrary values for conservative estimation of parameters. In addition, since usually, value of α is small, around $2.5*10^{-4}$ (Femeena et al., 2019), logarithmic scaling of parameter space of α may help to find the optimal solution by taking more samples around assumed small values.

According to study of Szeftel et al. (2011) lateral water inflow structure significantly affect to estimated TSM parameters. Therefore it is important to clarify lateral inflow and outflow condition. In this study, when upstream discharge and downstream discharge are different, lateral inflow is uniformly assigned as averaged value throughout reach length as Eq. (3.14).

$$q_L = \frac{Q_{downstream} - Q_{downstream}}{L} \tag{3.14}$$

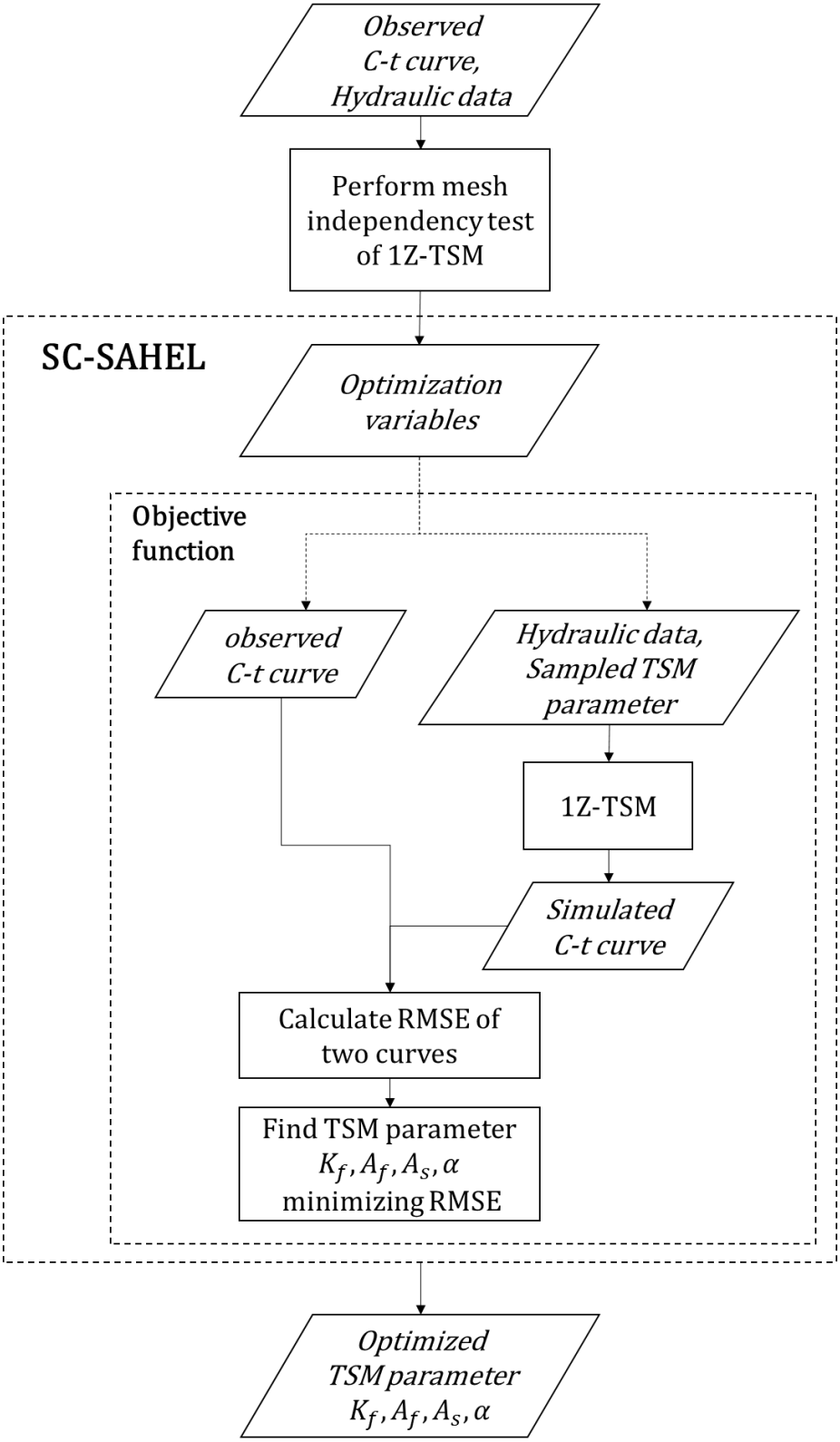


Figure 3.3 The flowchart of the TSM parameter estimation framework

3.3 Further Assessments for the Parameter Estimation Framework

3.3.1. Tracer Test Description

In order to assess effect of mesh grid size on estimation of TSM parameters, several sets of parameters were estimated. Benchmark experimental case is tracer injection test carried out on Cheong-Mi Creek, Korea (NIER, 2015). The tracer test was conducted in October, 2015. In 3,550 m of experimental site, Hyun-sa bridge is located and the sinuosity of the site is 1.26. Meandering bends and constructions in natural river intensify solute mixing by disturbing main flow structure (Jackson et al., 2013). By considering the facts, Cheong-mi Creek experiment was selected as benchmarking problem of TSM parameter estimation.

During the experiment, discharge was 2.26 cms. Hydromorphic parameters were measured at 10 cross sections. Mean velocity and mean depth were measured using StreamPro ADCP of RDI. Plan view of conducted tracer test is illustrated as Figure 3.4. In the section denoted as S#, both of hydromorphic parameters and concentration curves were measured. I.P is tracer injection point and cross-sections denoted as U# are observation point for only hydroporphic parameters.

As a tracer material, 0.2 kg of Rhodamine WT, which is conservative and easily detected flourscent dye, was injected. The measurement device was YSI-600OMS fluorometer that detects the intensity of light emitted from Rhodamine WT. Three of these devices were installed laterally uniform at all sites. The tracer test was designe for one-dimensional mixing, and the location of S1 was determined considering complete mixing length of Kilpatrick (1982) as follows.

$$L_0 = 0.1\left(\frac{1}{n}\right)^2 \frac{UW^2}{E_z} \tag{3.15}$$

where L_0 is the distance from the injection point for complete mixing on cross-section $[\mathrm{L}]$; n is the number of injection points in lateral direction; and E_z is lateral mixing coefficient $[\mathrm{L}^2/\mathrm{T}]$. In this case, the distance, where solute is completely mixed, was 300 m. Thus completely mixed BTCs are able to be observed at 940 m from I.P., the location of S1.

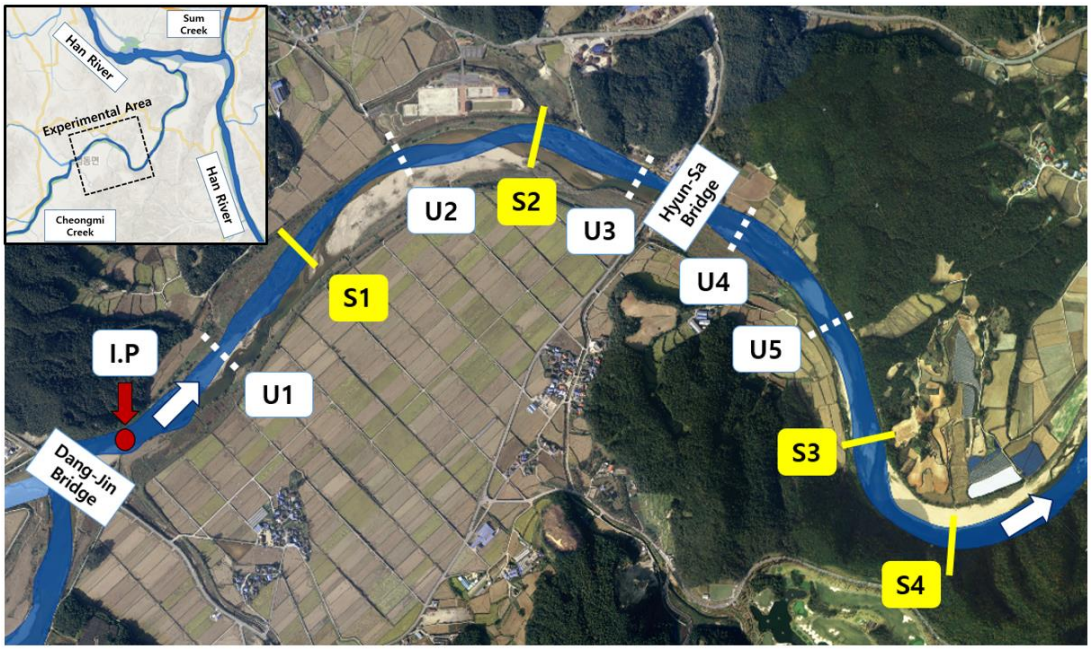


Figure 3.4 The plan view of Cheong-mi Creek experiment site (NIER, 2015)

Table 3.1 Measured hydraulic data in Cheong-mi Creek

Section	Distance from I.P. (m)	h (m)	W (m)	U (m/s)	U* (m/s)
I.P	0	0.72	17.1	0.19	0.023
U1	380	0.45	32.5	0.15	0.020
S1	940	0.33	17.5	0.39	0.055
U2	1,300	0.53	32.6	0.13	0.017
S2	1,690	0.63	31.7	0.11	0.014
U3	2,050	0.59	34.0	0.11	0.014
U4	2,410	0.35	16.5	0.39	0.055
U5	2,730	0.18	34.6	0.37	0.057
S3	3,080	0.39	14.1	0.41	0.056
S4	3,550	0.36	24.25	0.26	0.036
Average		0.42	25.48	0.210	0.028

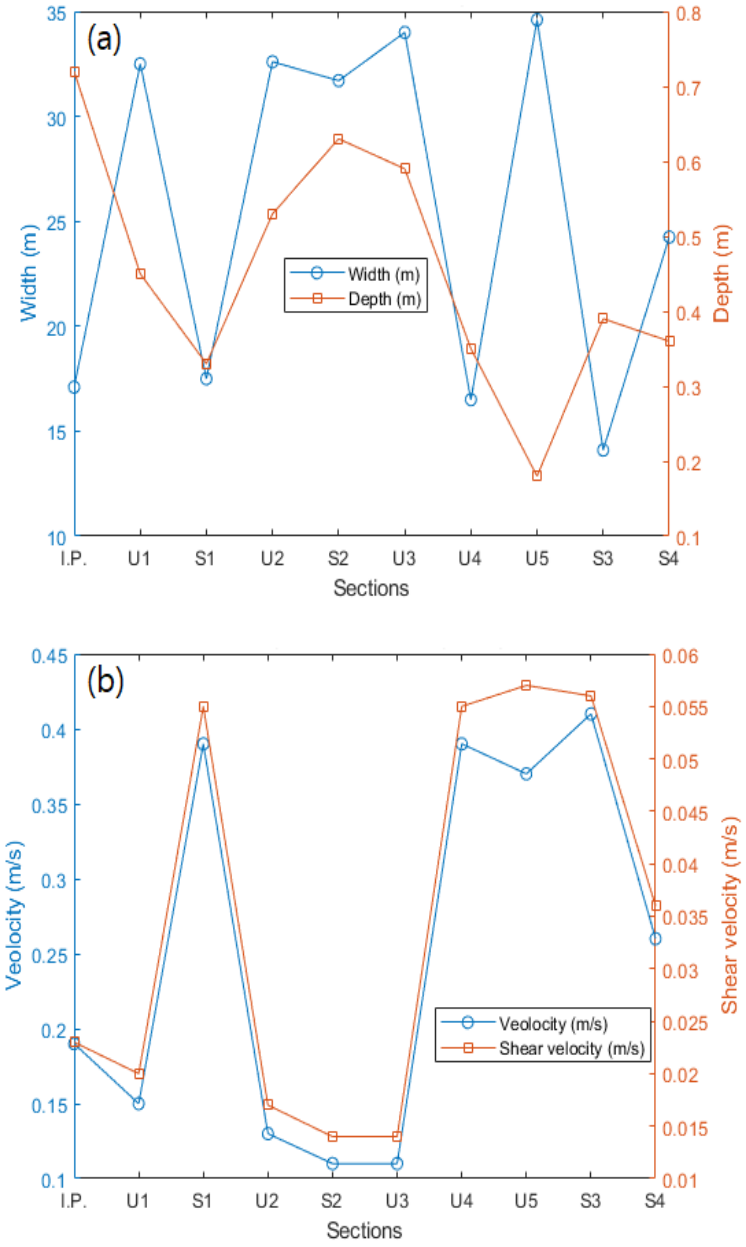


Figure 3.5 Plots of measured hydraulic data in Cheong-mi Creek;

(a) Width and depth; (b) Velocity and shear velocity

3.3.2. Grid Independency of Estimation

Since the number of segments is limited as under 750, OTIS may suffer from numerical inaccuracy apart from numerical stability guaranteed by Crank-Nicolson numerical scheme. In this manner, parameter estimation model of OTIS is subjected to numerical inaccuracy lead to false convergence. For instance, limitation of numerical accuracy may lead to wrong estimation value in large scale simulation, such as comparatively small value of dispersion coefficient which is influenced by numerical dispersion. Moreover, fundamental uncertainty in TSM parameter estimation, initial input values, and deficiency of measured data are associated with unreliable value (Kelleher et al., 2013; Femeena et al., 2019). For that reason, replica numerical model of OTIS was implemented to parameter estimation framework so that perform mesh independency test before parameter estimation for reliable estimation.

Used two concentration curves for grid independency assessment are measured in section 1 and 2 with reach length of 750 m. In these numerical experiments, the numbers of segments were varied from 16 to 201, and time step sizes were calculated at the same time (based on $CFL = 1$). The plan view and measured hydraulic values are given in Figure 3.4 and Table 3.1, respectively.

Parameter estimation results are shown in Table 3.2 and Figure 3.6. Especially, changed estimated values were also calculated in percent ratio of estimated value of 201 grids. The calculated change of the K_f is almost 2 0%, and A_s is 7 %. However, the results of A_f and α were insignificant compared to the other two parameters with under 1 % of differences. Since this effect comes from numerical truncation error, i.e. numerical dispersion, parameters accounting the dispersive shape of BTC, K_f and A_s, showed larger differences than others.

More discussing about numerical dispersion aspect, generally that numerical dispersion decreases longitudinal dispersion coefficient has been reported. However, in TSM, the dispersion coefficient increased contrary to the 1D-ADE case. It is because the compound effect of the dispersion coefficient and storage zone features modify BTC shape. In this manner, advective characteristic, A_f, was comparatively less affected. These differences of estimation values are decrease as grid size is denser as depicted in Figure 3.5. In this case, gradients were close to zero when the number of segments is over 75.

As discussed above, grid size has an effect on TSM parameter estimation, especially K_f and A_s. In order to obtain an unaffected simulation result, usually people who use numerical model conduct mesh independency test that finds optimal grid size by trial and error method. Taking effect of grid condition into account, the optimization framework, TSM-SC-SAHEL, in this study implemented mesh independency test for reliable parameter estimation. The mesh independency test is carried out decreasing grid size under assumed TSM parameters before estimation. Next, the grid size calculated from the test becomes a grid condition of parameter estimation. Accordingly, TSM-SC-SAHEL makes sure the estimation result is reliable value regarding numerical error adopting the test.

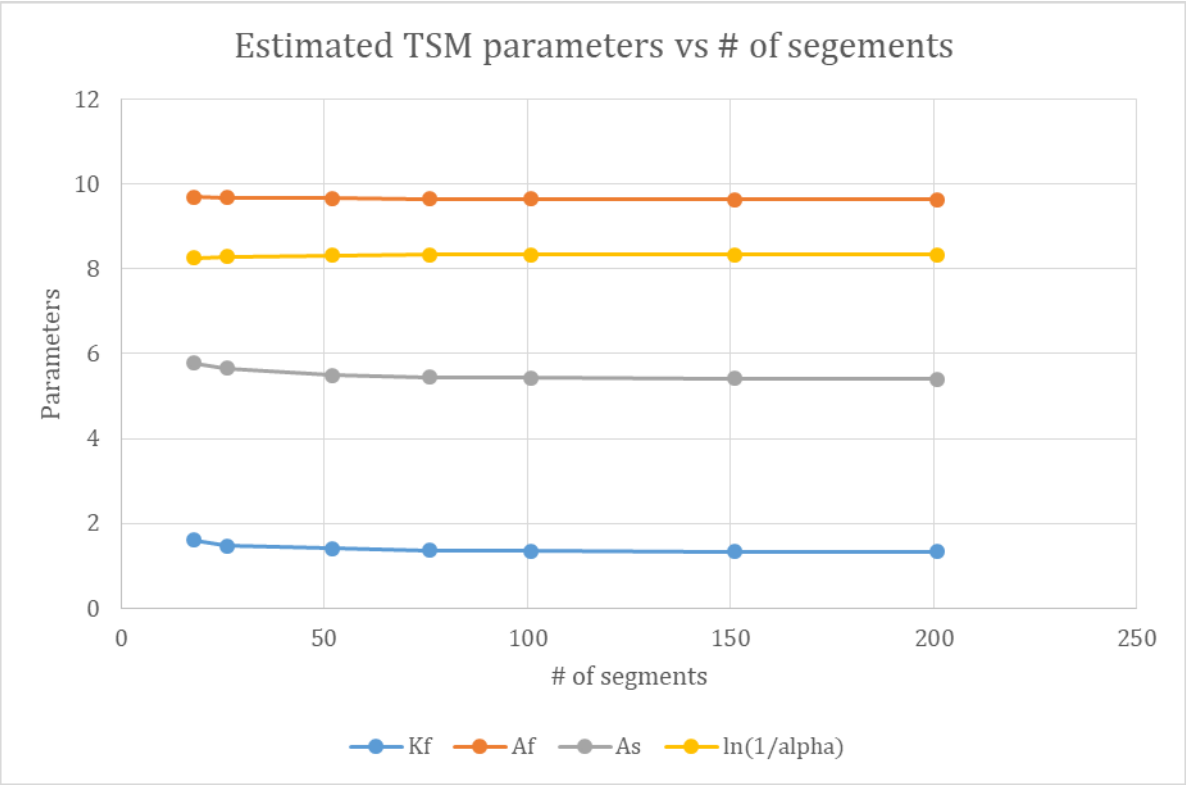


Figure 3.6 The plot of estimated TSM parameter values versus numbers of segments

Table 3.2 Estimated TSM parameters and grid sizes

			TSM parameters			
case	**dx (m)**	**dt (s)**	K_f **(** m^2/s **)**	A_f **(** m^2 **)**	A_s **(** m^2 **)**	$\alpha * 10^4$ **(** $1/\mathrm{s}$ **)**
Cheong-mi Creek, Korea, Station1-2	3.7406	13.3838	1.3353	9.6316	5.3971	2.4029
	4.9834	17.8302	1.3263	9.6315	5.4098	2.4109
	7.4627	26.7009	1.3436	9.6408	5.4284	2.4161
	9.9338	35.5423	1.3649	9.6508	5.4496	2.4218
	14.5631	52.1057	1.4017	9.6661	5.4928	2.4383
	29.4118	105.2331	1.4633	9.6829	5.6549	2.5191
	$\frac{\max(\Delta\text{parameter})}{\text{parameter}_{\#seg=200}}$ (%)		20.4160	0.5643	7.0355	0.9369

3.3.3. Choice of Optimization Setting

A number of researchers have been working on developing evolutionary algorithm to solve variety of hydrologic and environmental problems. However, researches keep reporting that every EA has pros and cons. For example, SCE-UA (Duan et al., 1993) suffers from a problem that population degeneration problem (Chu et al., 2011) even though this popular method has been regarded as reliable and robust. Due to such pros and cons of each algorithm, there is no algorithm that is consistently superior to other methods over various problems, and this is called as No-Free-Lunch (NFL) theorem of Wolpert and Macready (1997). Overcoming efforts for NFL theorem led researchers to develop a lot of adaptive algorithms such as A Multialgorithm Genetically Adaptive Method for Single Objective Optimization (AMALGAM-SO) (Vrugt and Robinson, 2007; Vrugt et al., 2009) and SC-SAHEL.

Since a question that, which EA is the best algorithm for a certain problem, is not easily answered, compatibility of EAs for TSM parameter calibration was checked in this section. For the assessment of compatibility, three fitness criterions (Mean Squared Error (MSE), Maximum Error (ME), Mean Absolute Error (MSE), and $1\text{-}R^2$) and the evolution numbers were compared for model fitness and optimization efficiency, respectively. Since concentration value of tail of BTC is generally small, every objective function was scaled as percent value by dividing peak concentration value C_{max}. Parameters were calibrated in case of Cheong-mi Creek, Korea (NIER, 2015). A total number of complexes was fixed as 10 for every EA settings in order to consider the evolution number as an index of convergence capability. The SC-SAHEL provides several popular EAs such as CCE of SCE-UA, and MCCE of SP-UCI. In other words, the SC-SAHEL framework make assessment, optimization condition comparison, easy. Hence, the SC-SAHEL optimization framework was adopted for comparison in this study. Nominated EAs

are CCE of SCE-UA, MCCE of SP-UCI, DEF, FL, GWO and combination of the five EAs supported in SC-SAHEL optimization framework.

The SC-SAHEL framework provides uniform distribution random sampling and Latin Hypercube Sampling (LHS) methods for initial population saplming step. In order to exclude effect of initial sampling start spot, the LHS was adopted as sampling method. Size of sample population and optimization settings of this study are arranged at Table 3.3. Total number of objective function evaluation is calculated as poduct of the number of complexes and the size of complexes. In such manner, the number of function evaluation in each evolution is 90 because each step has 9 individuals of 10 complexes.

In order to avoid case specificity of optimization method assessment as possible, every accessable BTC measured in Cheong-mi Creek was applied to parameter estimation. Thus, every 6 reach combinations of 4 cross-sections. Taking uncertatinty of parameter estimation into consideration, each optimization condition was iterated for 3 times. Particularly, 24 optimization conditions were compared under combinations of 6 EA conditions, and 4 objective functions. Therefore, 432 times of TSM parameter estimations were conducted for assessment. Detailed parameter estimation results of each reach and condition were given in Appendix. I.

In order to compare convergence of each optimization conidition, number of successful parameter estimation and success ratio were calculated as Table 3.4. In the table, success ratio of DE was 93 %, and it was the best among assessed method. Success ratio of the SC-SAHEL and the FL showed as 88 % and 86, % respectively. In the perspective of objective function, the PMSE showed highest success ratio 99 % and the Coefficient of Determination (CoD) showed following performance.

Mean evolution number, which represents convergence speed, was calculated as Table 3.5.

In order to assess efficiency of optimization condition, an Expected Evolution Number (EEN), that can be calculated as following equation (3.16), was also given.

$$\text{Expected Evolution Number (EEN)} = \frac{\text{Evolution Number}}{\text{Success rate}} \tag{3.16}$$

The EEN is needed evolution number for one successful parameter estimation. Total evolution number and EEN were 69, and 86, respectively. Most efficient method, which is low cost, was the MCCE with 46 times. For every objective function, best EA was the SC-SAHEL, that has 59 EEN, and it was 1.45 times faster than total EEN that is 86 times. In the objective function point of view, the CoD and the PMSE were fastest method showing 56 ~ 57 times of evolution number. Both CoD and PMSE has 58 times of EEN, that is 3 times faster than PMAE which showed 200 times of EEN.

Taking both evolution number and identifiability of global solution into account, even though DEF has an advantage in identifiability, the DE is not desirable, since it is the slowest method. However, in case of taking fitness function as MSE, which gives successful result in most methods, SC-SAHEL with all EAs gives the fastest and reliable solution. Consequently, SC-SAHEL method with the five EAs and the percent mean square error were adopted as a default setting of TSM parameter estimation framework of this work since this combination satisfies both robustness and convergence speed.

Table 3.3 Optimization parameter settings

Optimization Parameter	Setting
Total number of complexes	10
Size of complexes	9 (2 * dimension + 1)
Exploited EAs	CCE, MCCE, FL, DEF, GWO
maxn	10^{d}(default)
StopStep	50 (default)
StopIMP	0.1 (default)
StopSP	10^{-7}(default)

Table 3.4 Number and ratio of successful parameter estimations

Objective function		SCE-UA (SC-CCE)	SP-UCI (SC-MCCE)	SC-MFL	SC-DEF	SC-GWO	SC-SAHEL (CCE+MCCE +FL+DEF+GWO)	total
PMSE	#	18	17	18	18	18	18	107
	rate	**100 %**	**94 %**	**100 %**	**100 %**	**100 %**	**100 %**	**99 %**
PME	#	16	15	17	18	13	17	96
	rate	**89 %**	**83 %**	**94 %**	**100 %**	**72 %**	**94 %**	**89 %**
1-R2	#	17	17	18	18	18	17	105
	rate	**94 %**	**94 %**	**100 %**	**100 %**	**100 %**	**94 %**	**97 %**
PMAE	#	0	1	9	13	6	11	40
	rate	**0 %**	**6 %**	**50 %**	**72 %**	**33 %**	**61 %**	**37 %**
total	#	51	50	62	67	55	63	348
	rate	**71 %**	**69 %**	**86 %**	**93 %**	**76 %**	**88 %**	**81 %**

Table 3.5 Mean evolution number of Cheong-mi Creek TSM parameter estimation (Exp#: expected number)

Objective Function		SCE-UA (SC-CCE)	SP-UCI (SC-MCCE)	SC-MFL	SC-DEF	SC-GWO	SC-SAHEL (CCE+MCCE+FL+DEF+GWO)	Obj. func. Averaged
PMSE	#	43	33	50	60	118	40	57
	Exp#	**E: 43**	**E: 35**	**E: 50**	**E: 60**	**E: 118**	**E: 40**	**E: 58**
PME	#	61	57	60	89	206	59	89
	Exp#	**E: 69**	**E: 68**	**E: 64**	**E: 89**	**E: 285**	**E: 62**	**E: 100**
1-R2	#	38	36	49	59	113	42	56
	Exp#	**E: 40**	**E: 38**	**E: 49**	**E: 59**	**E: 113**	**E: 44**	**E: 58**
PMAE	#	57	58	72	81	105	68	74
	Exp#	**E: inf**	**E: 1,044**	**E: 144**	**E: 112**	**E: 315**	**E: 111**	**E: 200**
EA ave.	#	50	46	58	72	136	52	69
	Exp#	**E: 71**	**E: 66**	**E: 67**	**E: 77**	**E: 178**	**E: 59**	**E: 86**

Chapter 4. Development of Formulae for Predicting TSM Parameter

4.1 Dimensional Analysis

The principal factors which influence the measurable characteristics of the transient storage exchange of solutes in natural streams can be categorized into three groups: fluid properties, hydraulic characteristics, and geometric characteristics of the stream channel. The fluid properties are its density and viscosity. The cross-sectional mean velocity, shear velocity, channel width, and depth of flow represent the hydraulic characteristics. The bedforms and sinuosity are geometric characteristics. Refer to Table 2.2, the key parameters of the transient storage zone model can be formulated as follows.

$$f_1(K_f, A_f, A_s, \alpha, \rho, \nu, g, U, U_*, h, W, S_n, L) = 0 \tag{4.1}$$

in which ρ is the fluid density $[\mathrm{M/L^3}]$.

For fully turbulent flow in rough open channels, e.g. natural streams, the dependence on the Reynolds number ($\mathrm{Re}=UR_h/\nu$) and the Froude number ($\mathrm{Fr}=U/\sqrt{gh}$) are negligible (Cheong et al., 2007; Sahay, 2012). Harvey and Waner (2000) showed that storage zone ratio and friction factor have a positive relationship (Figure 4.1), by analyzing tracer experiments performed on US streams. In general cases, friction slope and drag coefficient are not available in most data set, and it is not easy to measure the shear features. However, the shear velocity ($U_* = \sqrt{gHS}$) is related to shear force effects. By adopting shear velocity as an input variable, the shear

feature can be included in the equation. Moreover, since the shear velocity is a function of energy slope, the Froude number is implicitly included in the calculation.

According to Szeftel (2012)'s remark, that lateral inflow affects to estimated TSM parameter values, considering lateral inflow as input variable is reasonable. Measuring the exact distribution of lateral inflow is not only difficult but also not provided in most case of study. On the other hands, if mean flow velocity is calculated as Eq (4.2), the empirical equation can contain lateral inflow effect by adopting mean flow velocity.

$$\bar{U} = \frac{Q_{mean}}{A_f} = \frac{Q_{upstream} + Q_{downstream}}{2A_f} = \frac{2Q_{upstream} + q_L L}{2A_f} \tag{4.2}$$

Cheong et al. (2007) accounted that length effect on TSM parameter, using Peclet number. Since sinuosity S_n is calculated as ratio of reach length and distance ($= L / \text{Distance}$), length effect is also contained in both S_n and $\bar{U}$. Moreover, many investigators have been shown that dispersion effect have relationship with sinuosity. For instance, Jackson et al. (2013) classified that meander bend as STS taking relative effect of secondary flow and vortical structures, and according to Tonia and Buffington (2009), sinuosity causes hyporheic exchange. Thus sinuosity was nominated as third input variable of the equations.

By using the Buckingham's-Pi theorem, the following functional relationship between the various dimensionless terms is obtained.

$$\left(\frac{A_f}{Wh}, \frac{A_s}{Wh}, \frac{K_f}{hu*}, \frac{\alpha}{u*/h}\right) = f(W/h, U/U_*, S_n) \tag{4.3}$$

However, the mean cross-sectional area of the upper boundary and downstream boundary is not identical to the actual mean flow velocity. Regarding the facts that surface transient storage exchange procedure can be explained with sudden change of cross-sectional flow structure in general and hyporheic exchange is related to hydraulic conditions, mean advection proportion may not identical to the average of the upstream area and downstream area. In this context, main flow zone area was assumed to be an unknown parameter.

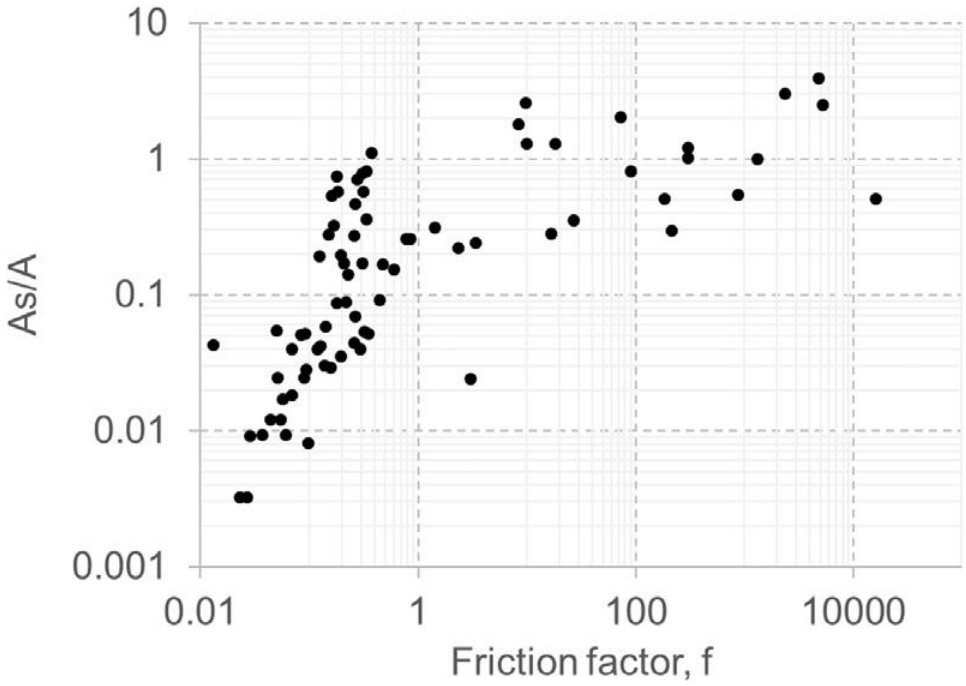


Figure 4.1 Scatter plot of storage zone ratio versus friction factor (Harvey and Wagner, 2009)

4.2 Data Collection via Meta-Analysis

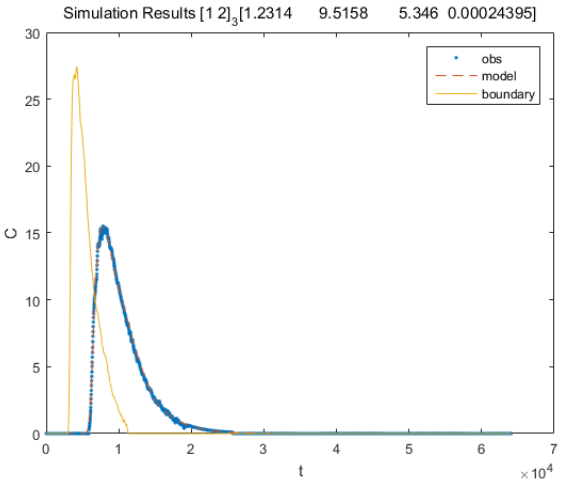
Gooseff et al. (2013) showed that sub-reach assignment has an effect on TSM parameter estimation. Furthermore, Rana et al. (2017) conducted an experimental study in the Jefferson National Forest in southwestern Virginia. From six times of experiment, measured lengths of two reaches were both 80 m, they showed that estimation of TSM parameter affected by the number of in-stream structures, and discharge difference. The given Table 4.1 is hydraulic data measured in the Cheong-mi Creek, Korea (NIER, 2015) and Table 4.1 is estimated TSM parameters of each section, respectively. The two reaches were both 750 m in length. The result of parameter estimation combining reaches in this study supports Gooseff et al. (2013)'s discovery. In Table 4.1, averaged parameter value of section 1-2 and 2-3 is different from the value parameter set of section 1-3. Taking into account these experimental studies that the estimated parameter set depends on how sub-reach condition is assigned, estimating TSM parameter set of every combination of the section is advantageous to obtain many and good quality of data set. Hence, estimations for meta-analysis were conducted in every possible condition where hydromorphic data set was acquired.

Even though OTIS-P has a local optima problem, estimated parameter sets by OTIS-P from published articles were assumed as global optimal value and the datasets were used in regression of empirical equations for four TSM parameters. In addition, Contrast to Femeena et al. (2019)'s study, shear velocity and sinuosity were adopted as input variable of equations in order to assess how much the variables affect to each TSM parameters.

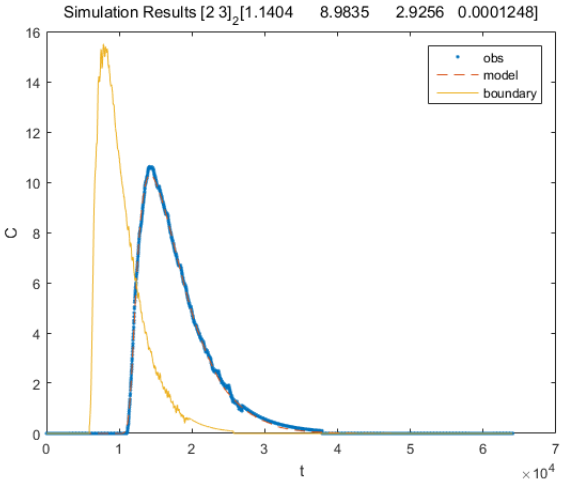
Since estimated 1D-ADE dispersion coefficient from certain case may contain storage zone effect, generally, the dispersion coefficients from 1D-ADE is larger than from TSM. Hence, estimated parameters based on both 1D-ADE and other storage zone models were not included,

Table 4.1 Estimated TSM parameters in from station 1 to 3

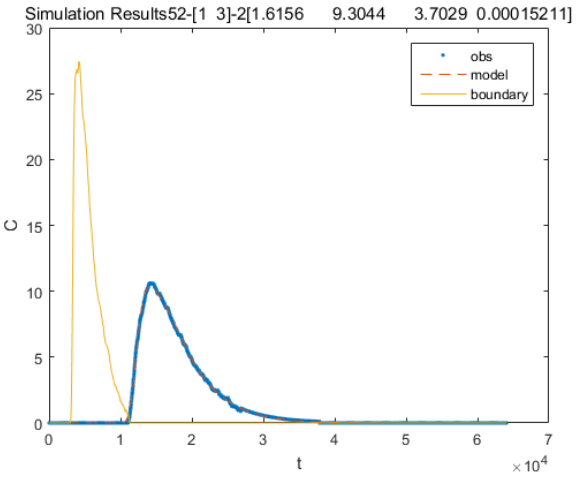
Study Area	Station	TSM parameters				
		K_f (m^2/s)	A_f (m^2)	A_s (m^2)	$\alpha * 10^4$ $(1/s)$	*DaI*
Cheong-mi Creek	1-2	1.3224	9.6304	5.4102	2.4118	2.143
	2-3	1.2028	9.0363	2.9521	1.2440	2.808
	avaraged of 1-2,2-3	1.2626	9.3334	4.18115	1.8279	2.475
	1-3	1.7023	9.3698	3.7284	1.5115	4.711



(a)



(b)



(c)

Figure 4.2 Plot of simulation result with estimated TSM parameters; (a) Station 1-2; (b) Station 2-3; (c) Station 1-3.

Table 4.2 Simple statistics of dimensionless TSM parameters [Cf: mean±1std (min ~ max)]

	TSM Parameters			
	$\frac{K_f}{hU_*}$	$\frac{A_f}{Wh}$	$\frac{A_s}{Wh}$	$\frac{\alpha}{U_*/h}*10^4$
Training (95 sets)	55.49±68.78 (1.129~306.3)	0.9450±0.2522 (0.4761~1.833)	0.2303±0.1504 (0.01171~0.8874)	13.64±14.32 (0.0467~57.80)
Testing (40 sets)	53.40±75.19 (1.342~249.7)	0.9066±0.1953 (0.5326~1.424)	0.2472±0.1874 (0.03717~0.8930)	15.87±16.60 (0.8683~53.640)

Table 4.3 Simple statistics of dimensionless input variables[Cf: mean±1std (min ~ max)]

	Input Variables		
	$\frac{U}{U_*}$	$\frac{W}{h}$	S_n
Training (95 sets)	3.252±2.589 (0.1746~10.54)	32.12±17.31 (5.75~87.14)	1.113±0.1535 (1~1.669)
Testing (40 sets)	3.447±2.461 (0.2614~8.044)	30.25±16.12 (11.4~75.56)	1.119±0.2094 (1~1.669)

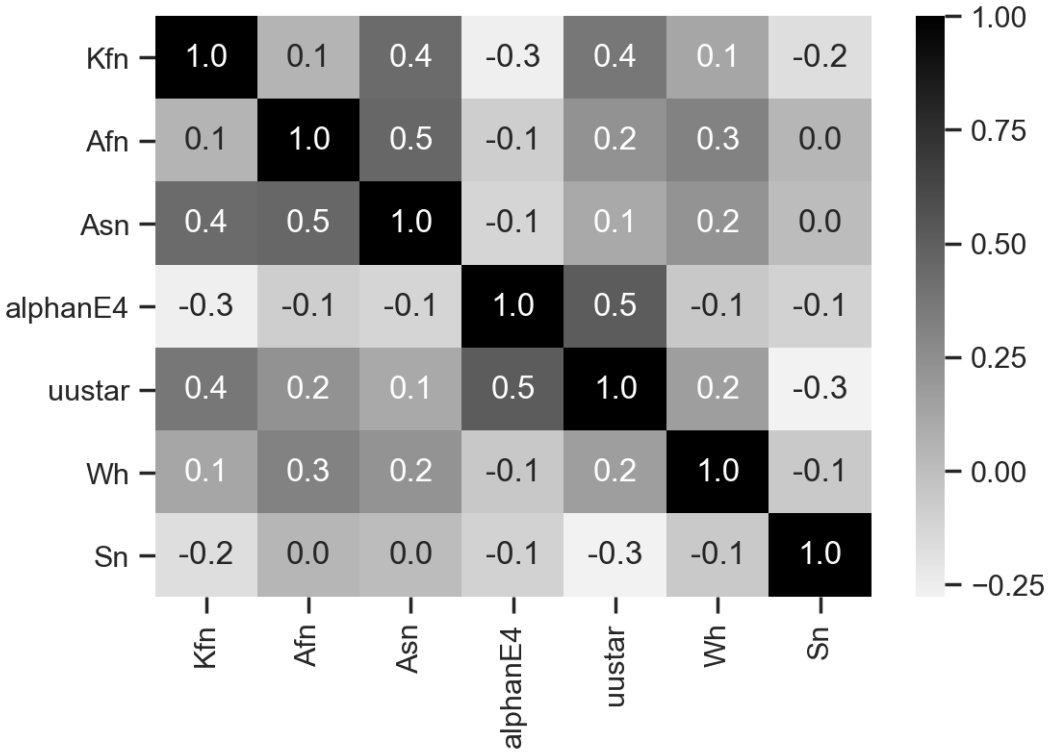


Figure 4.3 Pearson's correlation matrix of TSM parameters and input variables

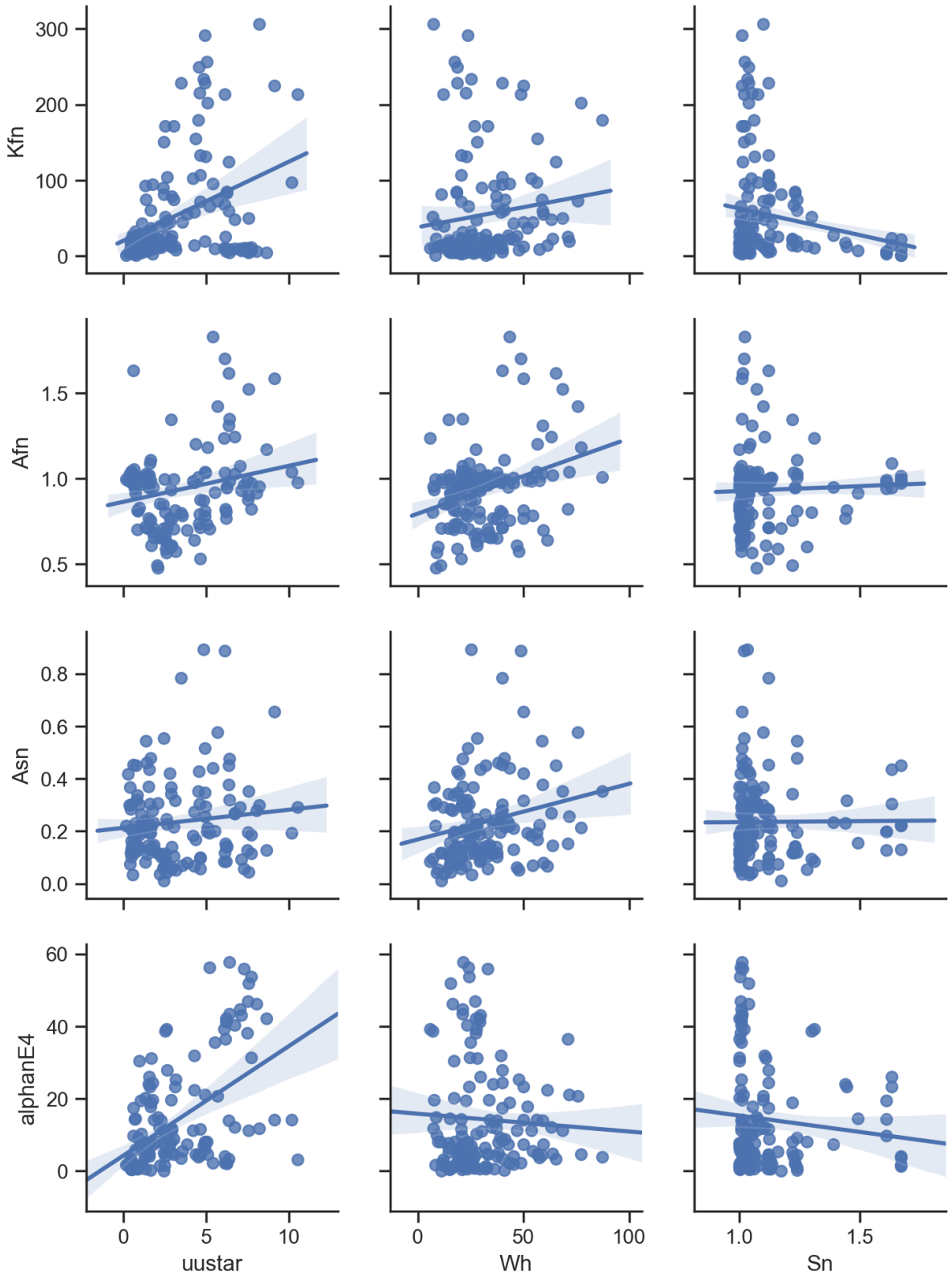


Figure 4.4 Pairplots of TSM parameters and input variables

unlike used data from Femeena et al. (2019). Instead of taking the estimated parameters based on other models, four TSM key parameters were re-estimated if BTCs were supported in the papers. In the aspect of injection type, tracer injection type (pulse injection and continuous injection) has an effect on estimated TSM parameter value (Wlostowski et al., 2013). Despite, by assuming the effect is not considerable, every TSM parameter produced from both of experiment types have been considered stream characteristic in this study to acquire abundant dataset.

Datasets were collected for meta-analysis from published literatures about tracer test in natural streams (Godfrey and Frederick, 1970; Graf, 1995; Czernuszenko et al., 1998; Harvey et al., 2003; Ensign and Doyle, 2005; Cheong et al., 2007; Gooseff, 2007; Rowiński et al., 2007; Bukaveckas, 2007; Stofleth et al., 2008; Gücker et al., 2009; Claessens et al., 2010a, 2010b; Stondahl et al., 2012; Johnson et al., 2014; NIER, 2015; Muller Price et al., 2016; Bohrman and Strauss, 2018; T. T. Le et el., 2018). A total number of the applicable set is 193, however, in order to develop robust equations, 25 % of datasets were assumed as an outlier that was excluded using a Robust Principal Components Analysis (RPCA; Hubert et al., 2005). Finally applied datasets in this study consist of 135 sets of complete TSM parameter combination in case of that information about nominated input variables are provided. The datasets were separated into training part and testing part. The training part includes 95 sets of parameter set with a range that cover total data, and the testing part contains 40 parameter sets. Among the datasets, the number of estimated parameter sets using the estimation method in this study are 96 sets. The simple statistics, mean, minimum, maximum, standard deviation values of TSM paramters and input variables are shown in Table 4.2-3. Detailed information about dimensionless variables and TSM parameters were given in the Appendix.II

Prior to the derivation of empirical equations, correlation of TSM parameters and input

variables was analyzed by calculating Pearson's correlation coefficient. According to Figure 4.3-4, the storage zone area A_s has a positive relationship with both A_f and K_f. In the relationship of the A_f and A_s, as the A_f increases as A_s increases. Surely, it is because these are subjected to stream of scale. Despite, A_f is correlated only storage zone because A_f, advection characteristic, does not contribute modification of BTC but composite transient storage process changes the shape of BTC. Moreover, relationships between K_f and TSM variables were observed. Especially, the correlation of dispersion coefficient and storage zone area is 0.4. Relationships of the K_f and storage parameters, the α and A_s, can be explained the effect of stream size on the exchange process. In a large stream, distance from river boundary to the center of flow zone is longer than small stream so that storage retention time, reciprocal to α, decreases. Meanwhile, investigators have been showing that dispersion coefficient is proportional to stream size in general since the proposal of Elder (1959)'s empirical equation. In this regard, the K_f is proportional to the A_s and $1/\alpha$, since the K_f, A_s and $1/\alpha$ have positive relationships with stream size, respectively.

In the aspect of input variables, calculated result agrees with previous studies. The dispersion coefficient was correlated with shear and stream morphological characteristic which are exploited in many empirical equations (Cheong et al., 2007, Sahay, 2012). As aforementioned, free flow zone area and storage zone area are subjected to the W/h. Moreover, the U/U_* is related to the variables. Especially, because the A_f means advection property, the ratio of advection and shear stress, the U/U_*, is related to the parameter indubitably. On the other hand, channel sinuosity is negatively correlated to dispersion and exchange rate due to secondary flow and head gradient. According to Jackson et al. (2013), low sinuosity, small

bend radius, forming small point bar decreases mean retention time. In other words, meander bend increasing residence time. In the dispersion point of view, an increase of retention time means a decrease of travel time in free flow zone that the time of solute transport is influenced by dispersion. This interpretation also describes the negative relation of the U/U_* and S_n. In morphological point of view, the reason is that meander interrupts flow path and increases the shear mixing layer (Jackson et al., 2013).

In brief, K_f is related to other TSM parameters and input variables, except for free flow zone area, but there is no variable that has a significant relationship with every input variable at the same time except for α and U/U_*. The result of correlation analysis between variables implies that TSM parameters have non-linear expressions.

In addition, the multicollinearity of input variables was assessed. In a regression problem, an interaction effect between independent variables influences on quality of regressed model. Especially, it is known that multicollinearity causes a problem in estimation and analysis (Belsley et al., 1980). For example, the Variance inflation factor (VIF) (Eq. 4.4) can be used to estimate multicollinearity. If VIF of the independent variable is less than 10, it is regarded that the independent variable has small multicollinearity. The VIF can be calculated as follows:

$$\mathrm{VIF}_i = \frac{\sigma^2}{(n-1)\mathrm{Var}[X_i]} \frac{1}{1-R_i^2} \tag{4.4}$$

where R_i^2 is the i the variable's coefficient of determination. Table 4.5 is calculated VIF factors of input variables in this study. Three input variables are having own information

according to VIF calculation since values are under 10.

Table 4.4 Calcualted VIF values of input variables

Variable	VIF factor
U / U_*	2.046
W / h	2.069
S_n	2.499

4.3 Formulae Development

A GPTIPS is MGGP library for Matlab that has been developed by Searson (2010, 2015). The GPTIPS was used in this study for the development of TSM equations. The fitness function setting was RMSE. Since best models are chosen under multi-objective both simplicity and best fitness value, a user has to maximum search space by tuning maximum depth and number of gene trees. Mathematical operators for symbolic regression should be specified in the execution of the MGGP. Because of reported non-linearity in TSM parameters, the hyperbolic tangent, and several exponentiate operators were considered in addition to basic arithmetic operators. Determination of the number of genes and the maximum tree depth indirectly means bounding maximum complexity of produced models. If finally presented model is too complicated to use, the model is not competitive than other models, especially than black-box models. Therefore, the limitation of the complexity in performing regression was varied from 4 to 15. In order to prevent the MGGP circulating around the certain structure, a fairly large number of regression processes was executed iteratively by setting termination time and adopting the recommended ratio of genetic evolution operators. Detailed parameter settings of the MGGP in this study are shown in Table 4.6.

Since the models what MGGP produces have various structure, equations with $R^2 > 0.5$ for dimensionless form in both testing and test set were nominated as the final result. Among nominated formulae, equations whose low complexity were presented below. The formulation based on MGGP is given as follows

$$\frac{K_f}{hU_*} = 1800\exp(\sqrt{\frac{U}{U_*}}) - 2011\frac{U}{U_*} + 1866\exp(-\frac{U}{U_*}) \\ -18.4\exp(\exp(\frac{U}{U_*} - \frac{W}{h})) + 2300\exp(-\exp(-\exp(\frac{U}{U_*}))) \\ -0.0622\tanh(\frac{W}{h} - \frac{U}{U_*})\exp(\frac{U}{U_*}) - 35.3(\frac{U}{U_*})^3 + 1.78(\frac{U}{U_*})^4 - 5633 \tag{4.5}$$

$$\frac{A_f}{Wh} = 0.281(S_n + \frac{W}{h} + 1.49) - 67.2\frac{U}{U_*} + 148\exp(-S_n) + 1.5*10^4\exp(-\frac{W}{h}) \\ -\frac{10.9U/U_*}{S_n^{W/h}} + 9.51(S_n + \frac{U}{U_*})^2 - 0.531(\frac{U}{U_*})^3 + \frac{14.31U/U_*}{S_n^{W/h}} + 37 \tag{4.6}$$

$$\frac{A_s}{Wh} = 15550\exp(-2S_n^{W/h}) - 4.18*10^5\exp(S_n^{W/h}) - 18\frac{W}{h} + 10440\exp(-S_n^{3W/h}) \\ +4.2*10^5\tanh(\exp(-S_n^{W/h})) + 25.3\tanh(S_n^{W/h})\tanh(S_n)(S_n + \frac{W}{h} + \sqrt{\frac{W}{h}}) - 24.5 \tag{4.7}$$

$$\frac{\alpha}{U_*/h}*10^4 = 0.0919\frac{W}{h} + 1.1*10^4\exp(-\frac{W}{h}) + 26.5\tanh(\frac{U}{U_*})^{(\frac{W}{h})^3} \\ -22.1(\frac{U}{U_*})^{3S_n}\tanh(\frac{U}{U_*}) + 9.03\frac{U}{U_*}\exp(-S_n^{W/h}) + 22.1(\frac{U}{U_*})^{3S_n} + 0.949 \tag{4.8}$$

Table 4.5 Parameter settings for the MGGP

Parameter	Settings
Function set	$\times, \div, +, -, \sqrt{}$, square, cube, exp, tanh, power
Population size	300-600
Number of generations	500-1000
Maximum number of genes allowed in an individual	4-15
Maximum tree depth	2-6
Tournament size	15
Elitism	10 % of population
Crossover events	0.84
High level crossover	0.2
Low level crossover	0.8
Mutation events	0.14
Sub-tree mutation	0.9
Replacing input terminal with another random terminal	0.05
Gaussian perturbation of randomly selected constant	0.05
Direct reproduction	0.05

In addition to the MGGP based equations, the Classical Principal Components Regression (CPCR) based equations were derived as Eq (4.9-12). Because outliers of raw data were excluded using RPCA, the CPCR method was adopted than the robust PCR method. The number of principal components used in regression was 3, identical to the number of input variables.

$$\frac{K_f}{hU_*} = \exp(1.9751)\left(\frac{U}{U_*}\right)^{0.5839}\left(\frac{W}{h}\right)^{0.2788}\left(S_n\right)^{-0.4465} \tag{4.9}$$

$$\frac{A_f}{Wh} = \exp(-0.5101)\left(\frac{U}{U_*}\right)^{-0.0035}\left(\frac{W}{h}\right)^{0.1236}\left(S_n\right)^{0.1418} \tag{4.10}$$

$$\frac{A_s}{Wh} = \exp(-3.1141)\left(\frac{U}{U_*}\right)^{-0.0593}\left(\frac{W}{h}\right)^{0.4234}\left(S_n\right)^{0.6635} \tag{4.11}$$

$$\frac{\alpha}{U_*/h} = \exp(-8.2818)\left(\frac{U}{U_*}\right)^{0.7338}\left(\frac{W}{h}\right)^{0.0786}\left(S_n\right)^{1.6476} \tag{4.12}$$

Chapter 5. Result and Discussion

5.1 Model Performances

After deriving new equations, five performance criterions, Root Mean Squared Error (RMSE), Discrepancy Ratio (DR), the accuracy based on DR (in a range of -0.3<DR<0.3), Pearson's correlation coefficient (ρ_{corr}), and coefficient of determinant (R^2); (Eq 5.1-4)) of Eq. (4.5-8), Eq. (4.9-12), and Femeena et al. (2019)'s equations were evaluated. $P_{(observed)}$ and $P_{(predicted)}$ are the observed and predicted parameter value matrix, respectively. The free flow zone area A_f was not calculated for Femeena et al. (2019)'s model since they did not propose equation for the A_f. Summary of performances in testing and training data is given in Table 5.1. In addition, the scatter plot of determined values and DR values are graphically presented in Figure 5.1-2 and 5.3, respectively.

$$\text{Accuracy (\%)} = \frac{100[\text{frequency}(-0.3 \le DR \le 0.3)]}{n} \tag{5.1}$$

$$DR = \log \frac{P_{(observed)}}{P_{(predicted)}} \tag{5.2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (P_{i(observed)} - P_{i(predicted)})^2}{n}} \tag{5.3}$$

$$\rho_{corr} = \frac{\text{COV}(P_{i(observed)}, P_{i(predicted)})}{\sqrt{\text{Var}(P_{i(observed)})\text{Var}(P_{i(predicted)})}} \tag{5.4}$$

$$R^2 = 1 - \frac{SSE}{SST} \tag{5.5}$$

First of all, DR based accuracy was evaluated. This accuracy criterion means how many close values were predicted. Accuracy values of Eq. (4.5-8) for each TSM parameter for the training set were 37.5 %, 100 %, 67.5%, and 52.5 %, for four TSM parameters respectively. In Eq. (2.48-50), 32.5 %, 45 %, and 27.5 % of accuracy were observed for K_f, A_s, and α, respectively. In terms of accuracy result Eq. (4.5-8) predicted better than Eq. (2.48-50) in every TSM parameters and The Eq. (4.9-12) showed intermediate performance of Femeena's model and the MGGP model. Furthermore, DR histograms were illustrated in Figure 5.3. As shown in the histograms and the scatter plots, large values of DR were observed in estimated parameters sets in Eq. (2.48-50). In the aspect of the K_f prediction using Eq. (4.5) is much accurate than using Eq. (2.48). Although Eq. (4.5) showed good performance in prediction of the K_f, prediction of the storage parameters, A_s and α, were slightly inferior in terms of RSME, ρ, and R^2. Especially, A_s prediction using Eq. (4.7) showed low performance compared to Eq. (2.49) except for accuracy. However, Eq. (4.8) showed better performance in ρ compared to Eq. (2.50) for α, but it gives unsatisfactory result in RMSE and R^2. For every criterion and parameters except for the A_f, equations based on PCR showed intermediate performance and performance was close to superior model than inferior model. Using both Eq. (4.6) and Eq. (4.10), calculated correlation coefficient and R^2 were larger than 0.9 in both training set and testing set. Prediction of the A_f of the PCR model in testing set was superior to the MGGP model while training set performance was inferior.

Predicted values and observed values are compared in Figure 5.1-3. In addition to Femeena

et al. (2019) and developed equations in this study, models of Pedersen (1977) and Cheong and Seo (2003) were compared in the K_f plots. From the result, Pedersen (1977)'s model overestimates the dispersion. Other three models showed fair result however models of Cheong and Seo (2003) and Femmena et al. (2019) showed overestimation compared to Eq. (4.5). Furthermore, from scatter plots of A_s and α, Eq. (4.7-8) showed accurate result while Femeena et al. (2019)'s equations underestimate A_s and α. As discussed above, The PCR models predict intermediate TSM parameter values.

Table 5.1 Calculated performances of empirical equations

Criteria	TSM parameter	This study MGGP		This study RPCR		Femeena et al. (2019) Eq. (2.48-50)	
		Training set (95 sets)	Testing set (40 sets)	Training set (95 sets)	Testing set (40 sets)	Training set (95 sets)	Testing set (40 sets)
Accuracy (%)	K_f (m^2/s)	**52.631**	**37.500**	49.474	30.000	42.105	32.500
	A_f (m^2)	**100**	**100**	**100**	**100**	-	-
	A_s (m^2)	**80.000**	**67.500**	72.632	60.000	51.578	45.000
	$\alpha * 10^4$ (1/s)	**46.315**	**52.500**	41.053	32.500	28.421	27.500
RMSE	K_f (m^2/s)	**11.015**	**7.988**	14.630	9.074	25.872	43.069
	A_f (m^2)	**4.991**	23.798	9.275	**12.230**	-	-
	A_s (m^2)	6.532	19.658	7.872	11.934	**5.008**	**10.864**
	$\alpha * 10^4$ (1/s)	6.778	6.694	**5.611**	3.971	7.281	**3.602**

Table 5.1 Calculated performances of empirical equations (continued)

Criteria	TSM parameter	This study MGGP		This study RPCR		Femeena et al. (2019) Eq. (2.48-50)	
		Training set (95 sets)	Testing set (40 sets)	Training set (95 sets)	Testing set (40 sets)	Training set (95 sets)	Testing set (40 sets)
ρ	K_f (m^2/s)	**0.925**	**0.838**	0.836	0.805	0.707	0.682
	A_f (m^2)	**0.994**	0.986	0.988	**0.992**	-	-
	A_s (m^2)	0.569	0.488	0.486	0.579	**0.581**	**0.784**
	$\alpha*10^4$ (1/s)	**0.527**	**0.494**	0.470	0.468	0.384	0.464
R^2	K_f (m^2/s)	**0.667**	**0.692**	0.414	0.603	-0.832	-7.951
	A_f (m^2)	**0.989**	0.931	0.963	**0.982**	-	-
	A_s (m^2)	-0.232	-1.441	-0.789	0.100	**0.275**	**0.254**
	$\alpha*10^4$ (1/s)	-0.214	-2.400	**0.168**	-0.196	-0.402	**0.015**

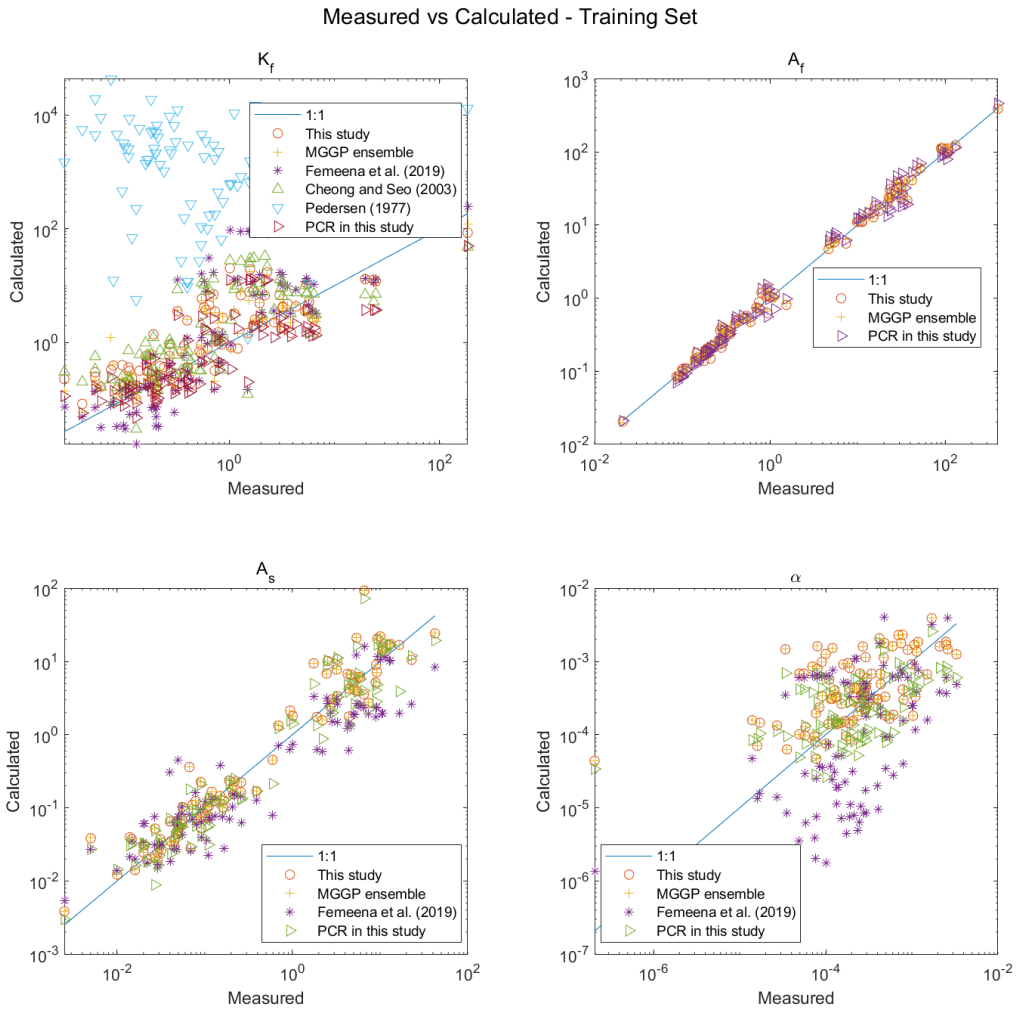


Figure 5.1 Comparison of determined and calibrated TSM parameters in training set

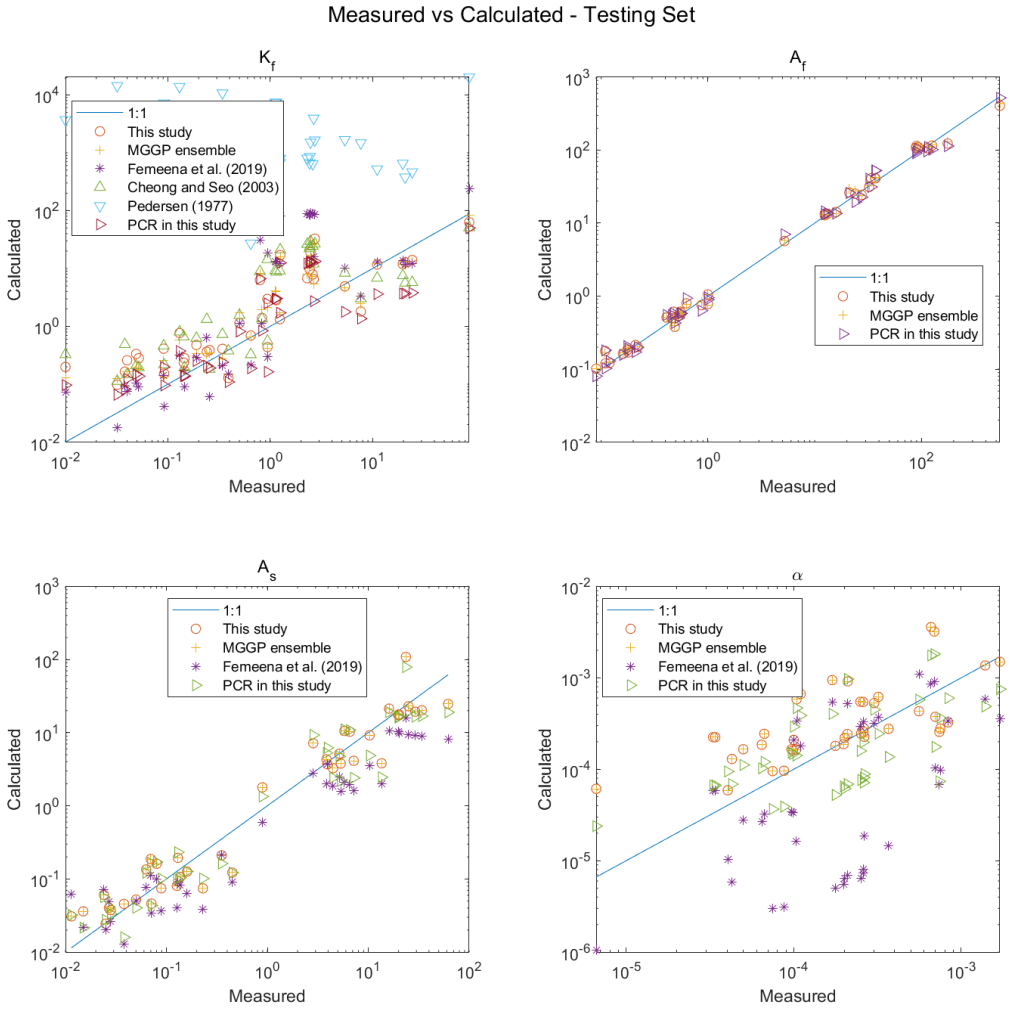


Figure 5.2 Comparison of determined and calibrated TSM parameters in testing set

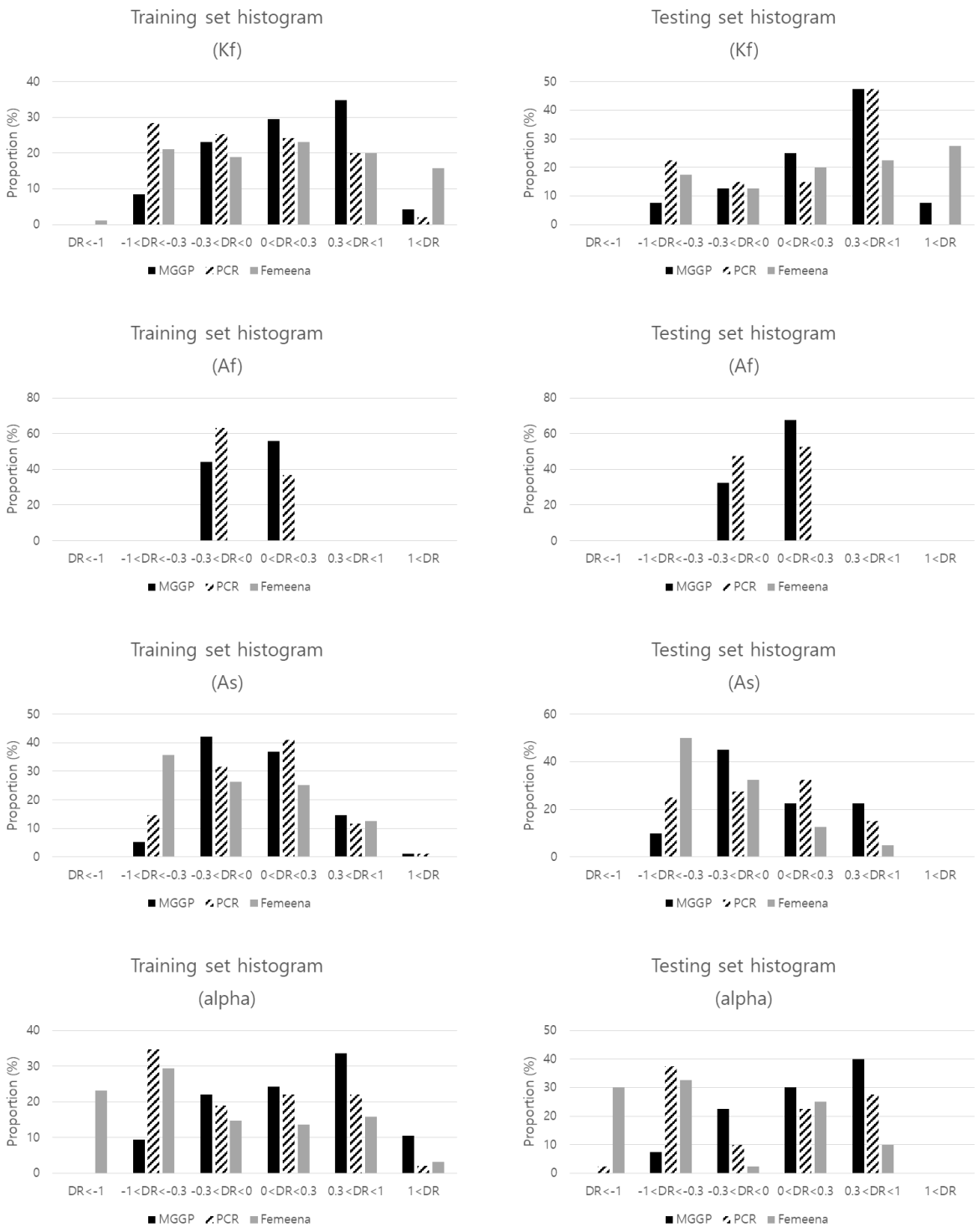


Figure 5.3 Discrepancy ratio plots of empirical equations for TSM parameters

5.2 Sensitivity Analysis

Even though developed TSM parameter equations using MGGP are complicated since the MGGP tries to describe the mechanism of the phenomenon by iteratively combining input variables. Hence, sensitivity analysis of developed empirical equations was conducted by One At a Time (OAT) plotting in order to understand TSM mixing mechanism relating river characteristics. Since proposed empirical equations were regressed in a range of training set. Thus, input variable sensitivity was analyzed based on plotting 20 % around the median value of the input variables. The sensitivity plots are given in Figure 5.4-5. However, the sinuosity was plotted from 1 to 3 since the S_n has an explicit range in natural stream (Figure 5.6).

The free flow zone dispersion coefficient K_f decreases as U/U_* increases. The Eq. (4. 5) has input variable W/h, but coefficients of terms which have W/h are small compared to other terms. Thus, terms with the W/h can be eliminated for the simplicity of equation if the target range is in the range of training dataset since the effect of the W/h is insignificant. This result reflects the discovery that longitudinal dispersion is proportional to the friction factor of stream. In previous studies about 1D-ADE, investigators showed that W/h effect is considerable. However, Figure 5.4 and Eq. (4.5) imply that the K_f have relatively small relationship with the W/h since TSM discriminates the effects of the dispersion in free flow zone and storage zone effect. In this respect, main flow property U/U_* is the most dominant factor among three dimensionless input variables for the dispersion coefficient of the TSM.

In terms of the A_f, which means an advective fraction, the U/U_* showed slightly negative relationship because velocity is reciprocal to flow section area in the same discharge. In

contrast, the W/h is proportional to the flow zone area. In general trend of the plot, the sinuosity of the stream increases the flow zone area. When flow meets meander bend, longitudinal momentum turns into the secondary flow. As a result of momentum change, meander bend decreases longitudinal velocity which is reciprocal to the A_f under the same discharge.

Calculated storage zone values by Eq. (4.7) are nothing on U/U_* in the equation structure point of view. However, the W/h and the S_n showed positive relationship according to Figure 5.5-6. In Figure 5.5, the W/h and the A_s showed proportional trend. The presumed reason is that STS, e.g. lateral cavity and pools, are more likely to exist in stream of large W/h. In a low sinuosity, 1~1.1, calculated parameters oscillated. It is presumed that it is because many cases of estimated TSM parameters on artificial experiment condition is usually on straight stream, e.g. restored agricultural stream. Nonetheless, the S_n and the A_s are proportional in general because sinuosity affects STSs and HTS (Tonia and Buffington, 2009; Jackson et al., 2013). As aforementioned, it is because secondary flow decreases the main flow zone fraction in flow structure.

The storage exchange rate α is the most complicated TSM parameter. Three input variables are correlated with the α according to equation form of Eq. (4.8) and Figure 5.4-6. The calculated result supports the physical interpretation of correlation discussed in section 4.2. In Figure 5.4, more relation is observed, generally, the U/U_* has a positive relationship, but gradient difference is observed in the change of both W/h and S_n. As discussed above, W/h is accounting for the existence of the STSs, and S_n is accounting for HTS. Moreover, S_n induces secondary flow transporting solute molecule to transient storages. This phenomenon is significant in large W/h and large S_n where streamflow fast, in large U/U_*.

Because in same morphological condition, high velocity makes secondary flow stronger so that storage exchange process more active.

In order to identify quantitative sensitivity values of variables around median value, spider plot of each parameter was illustrated in Figure 5.7. Moreover, two sensitivity indices, the elasticity (Eq. 5.6) and the local Sensitivity Index (SI, Eq. 5.7) of input variables, were calculated (Pannell, 1997).

$$e_{\text{var}} = \frac{\partial \text{parameter (\%)}}{\partial \text{variable (\%)}} \tag{5.6}$$

$$SI = \frac{\max(\text{parameter}) - \min(\text{parameter})}{\max(\text{parameter})} \tag{5.7}$$

As discussed above, the U/U_* is the dominant factor of the K_f. For the A_f , the U/U_* is dominant in both indices. The following order is S_n and W/h in both indices. However for a large sinuosity, the change of the A_f is larger than the change in a range of 80-120 % from median according to Figure 5.6. In terms of the A_s, the S_n and W/h are effective factors. The S_n is much sensitive variable than the W/h in this range. The exchange rate α is affected by every input variable, U/U_*, W/h, and S_n. Among the variables, the sinuosity was the most affecting variable, and the U/U_* followed.

Figures 8 – 10 are the sesnsitivity plots of the PCR model developed in this study. For the simplified view, spider plots were also depicted in Figure 11. Since equations of this model are having simple structure, sensitivity of variables are similar over range of each input variables. In other words, informations of Figures 8 – 10 are resembled in Figure 11. Thus, overall

sensitivity of input variables over TSM parameters can be assessed by analyzing Figure 11. Since the equations have power equation, absolute value elasticity and SI are having proportional relationship. Similar to the MGGP, the effect of U/U_* is most dominat factor to the K_f equation of the PCR model. Contrary to the MGGP, PCR model is affected by other factors, W/h and S_n. Elasticity of S_n was second effective factor that is negatively correlated with the K_f. As aforementioned, the reason of this sensitivity may be caused by secondary flow of channel meander. In spider plot of A_f showed different result in sensitivity of U/U_* and S_n. Most dominant factor in prediction of A_f was the sinuosity with elasticity of 0.035. Contrary to result of Figure 5.7, U/U_* has small effect to the free flow zone area. Since the storage zone are A_s is flow area property either, sensitivity of input varibles are similar to the free flow zone area A_f but portion of the U/U_* was larger in the A_s. In the aspect of the storage exchange rate α, the S_n is most correlated factor among the input variables. The W/h has almost no effect. The SI of W/h in α is 0.03 which is lowest value among calculated SI values.

Briefly, the dispersion effect on main flow zone is governed by friction ratio, U/U_* and it has negarive relationship with the S_n. In contrast, sensitive in the storage zone area A_s is sensitive to the W/h and it is positively related with the S_n. From presented figures, all input variables have significant relationships with both A_f and α but the W/h was weakly correlated with the storage parameters. Accordingly, in order to describe the TSM exchange mechanism properly, three input variables should be considered at the same time, especially the channel sinuosity.

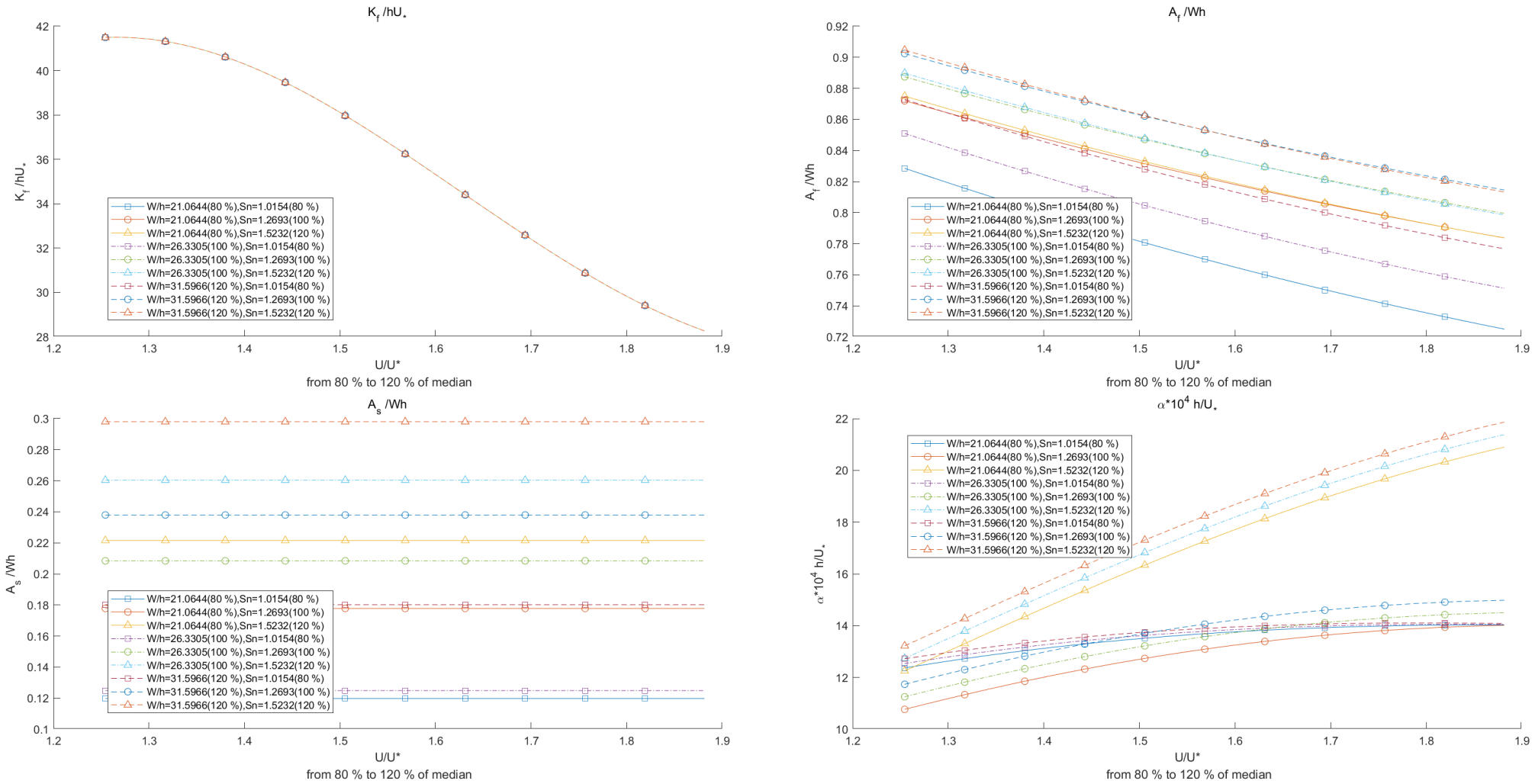


Figure 5.4 Calculated parameter values in a range of 20 % around the median of U/U_* (MGGP)

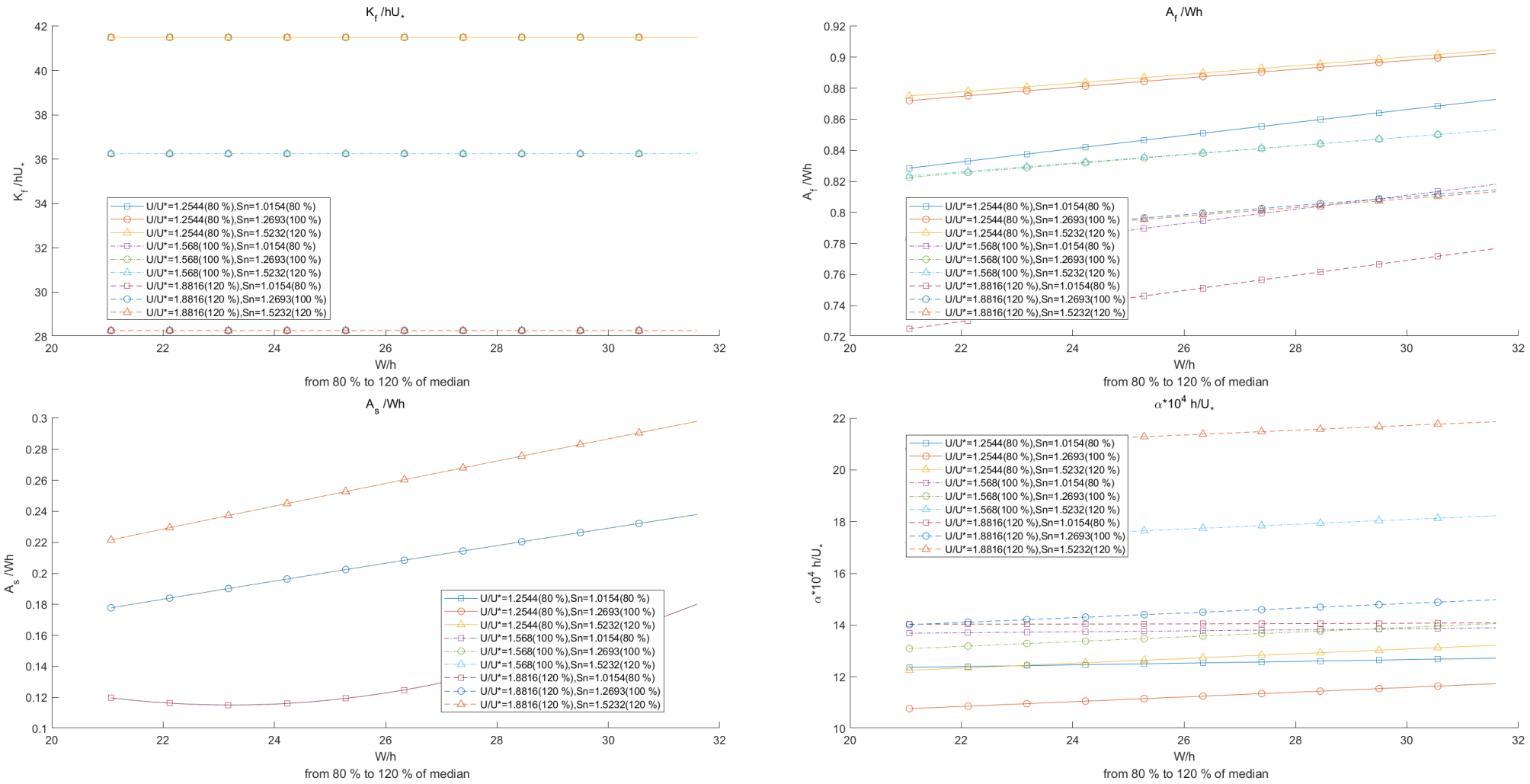


Figure 5.5 Calculated parameter values in a range of 20 % around the median of W/h (MGGP)

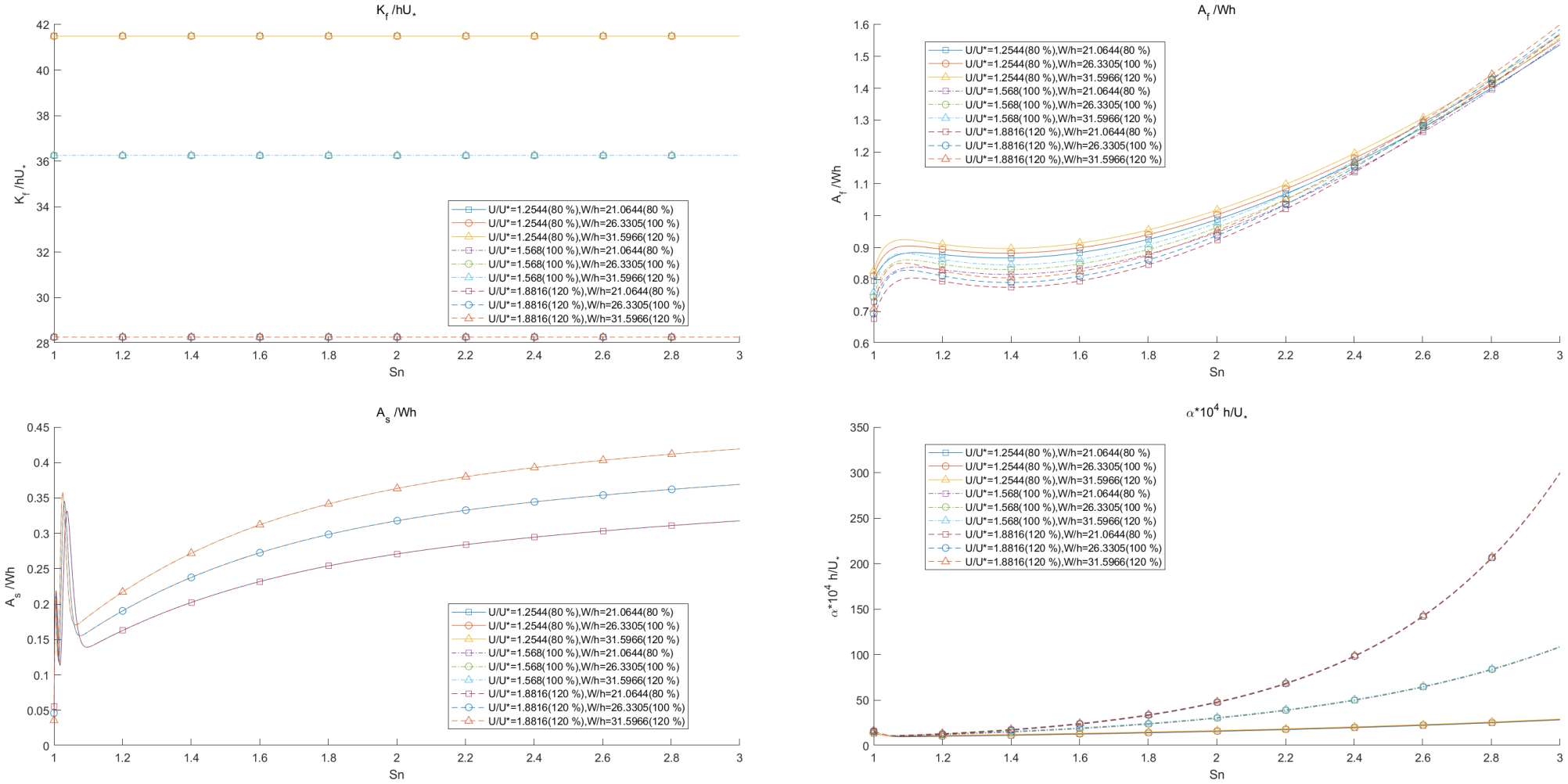


Figure 5.6 Calculated parameter values in a range of S_n (MGGP)

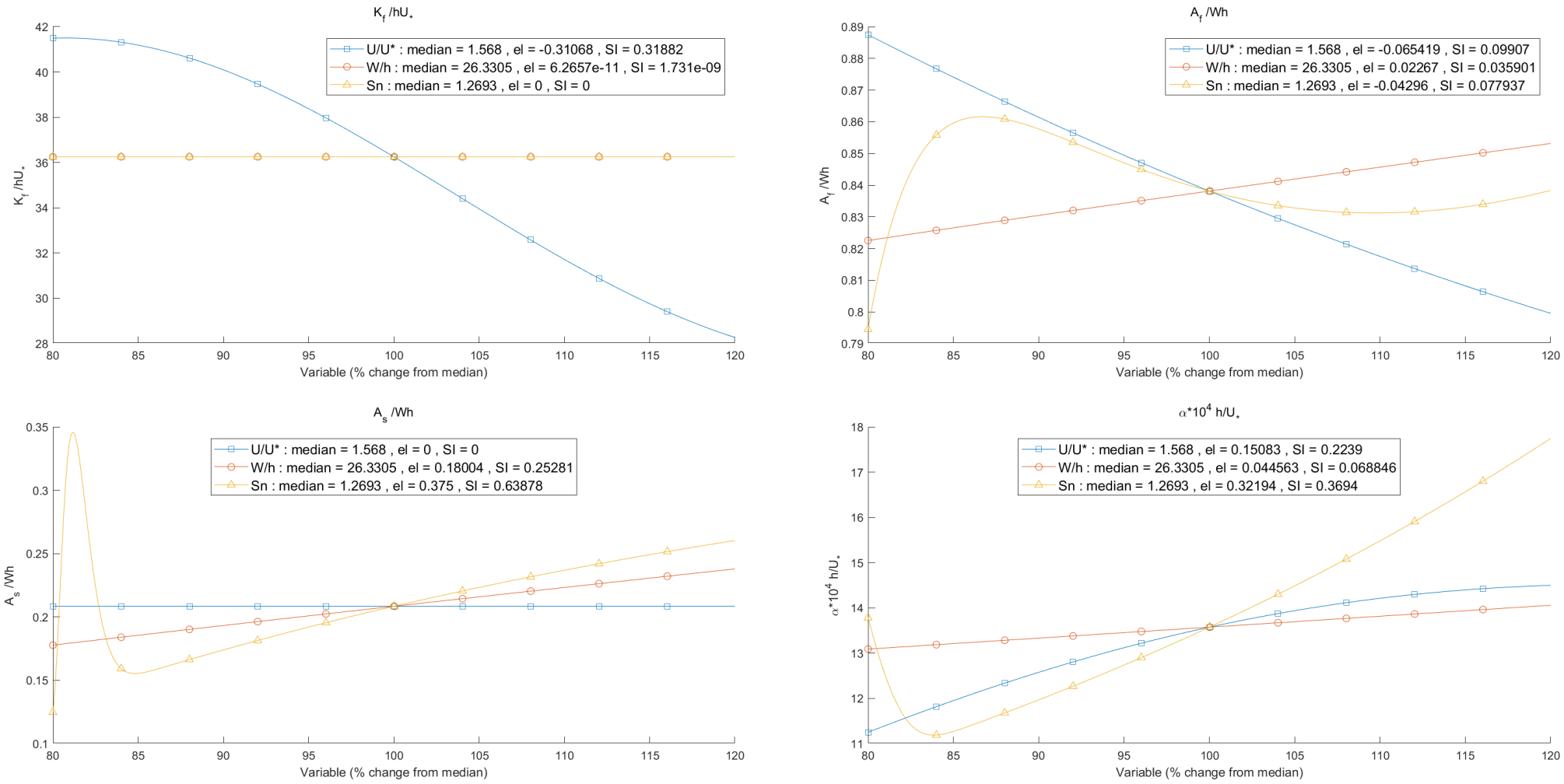


Figure 5.7 Spider graph of each TSM parameter and calculated sensitivity criterions (MGGP)

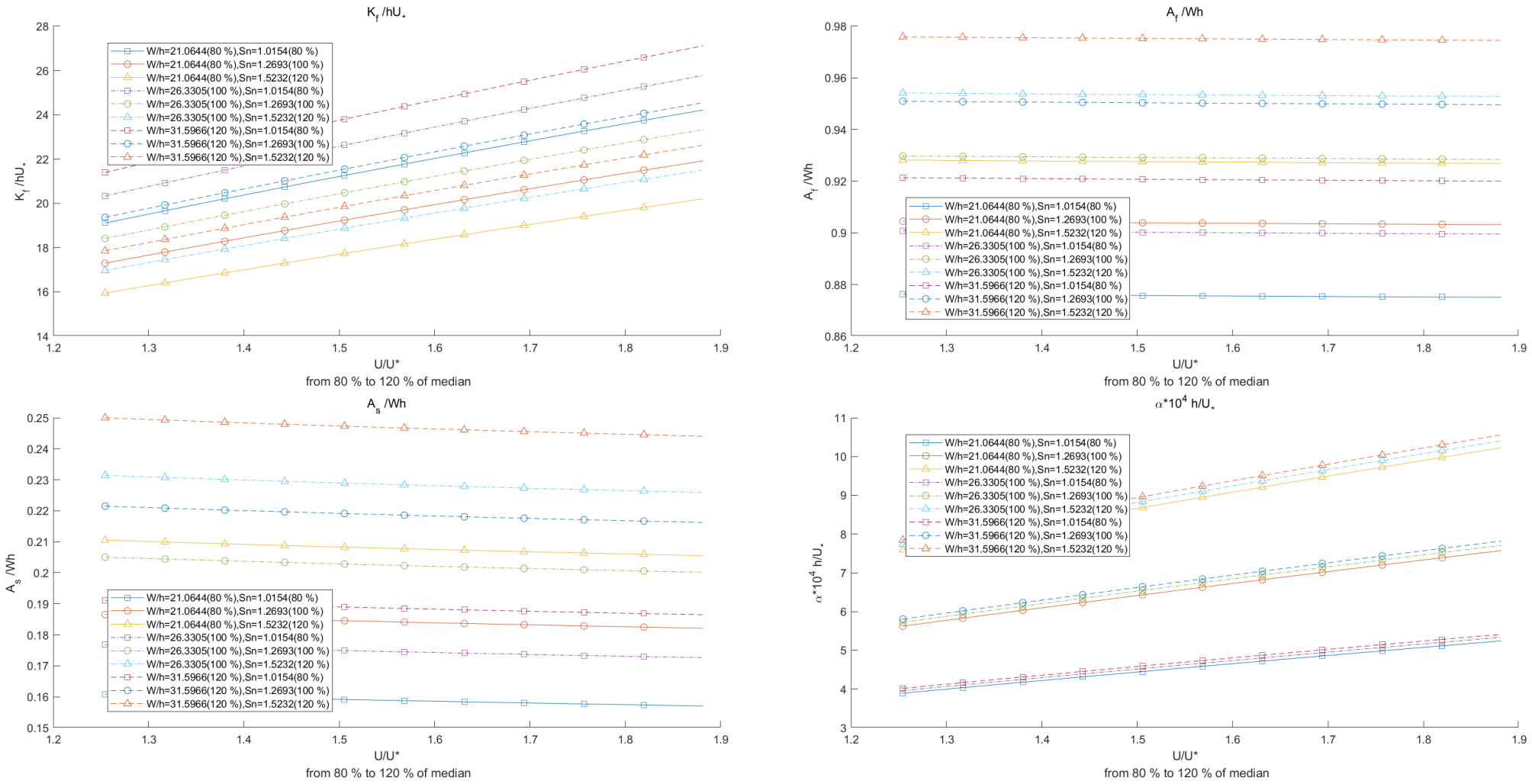


Figure 5.8 Calculated parameter values in a range of 20 % around the median of U/U_* (PCR)

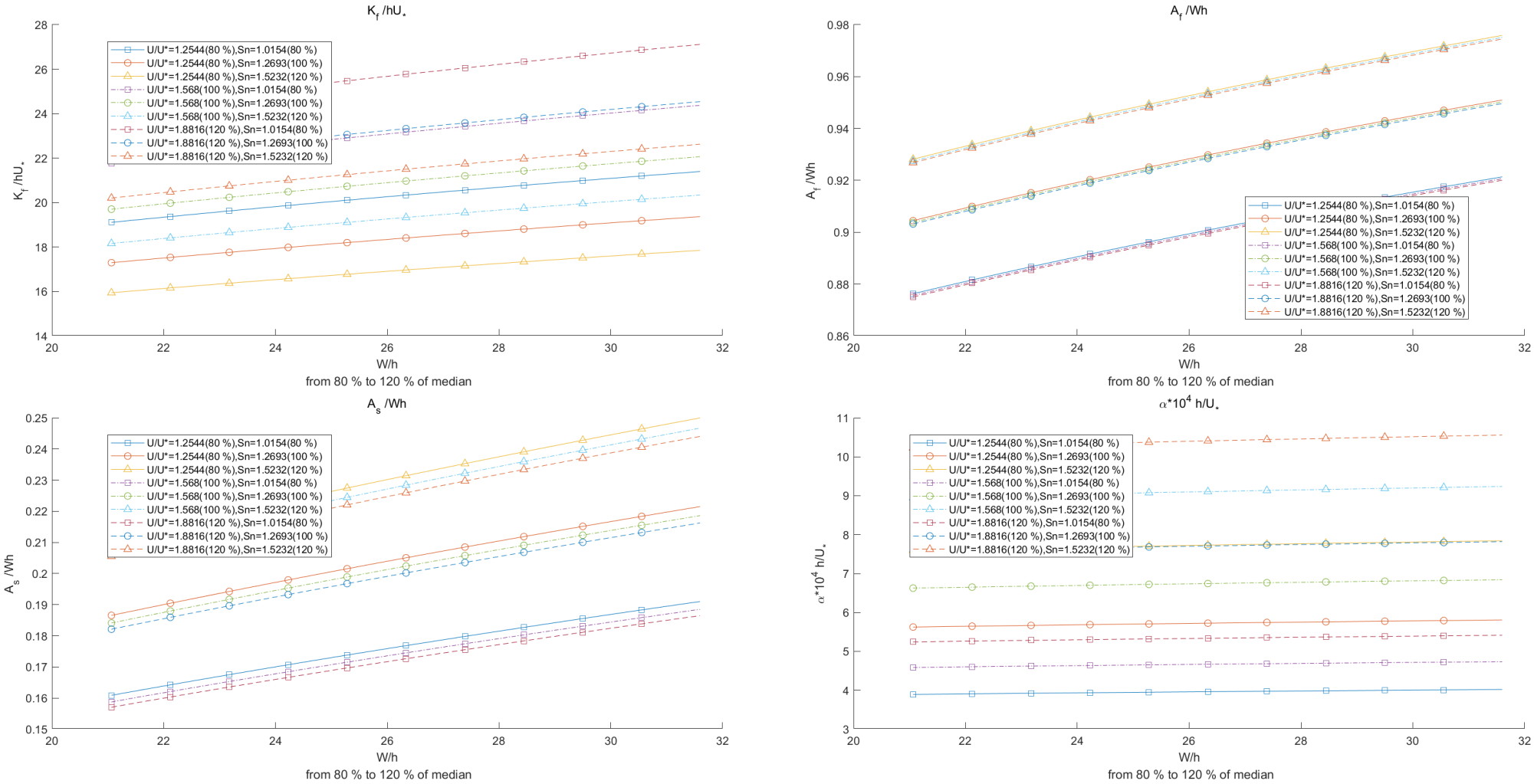


Figure 5.9 Calculated parameter values in a range of 20 % around the median of W/h (PCR)

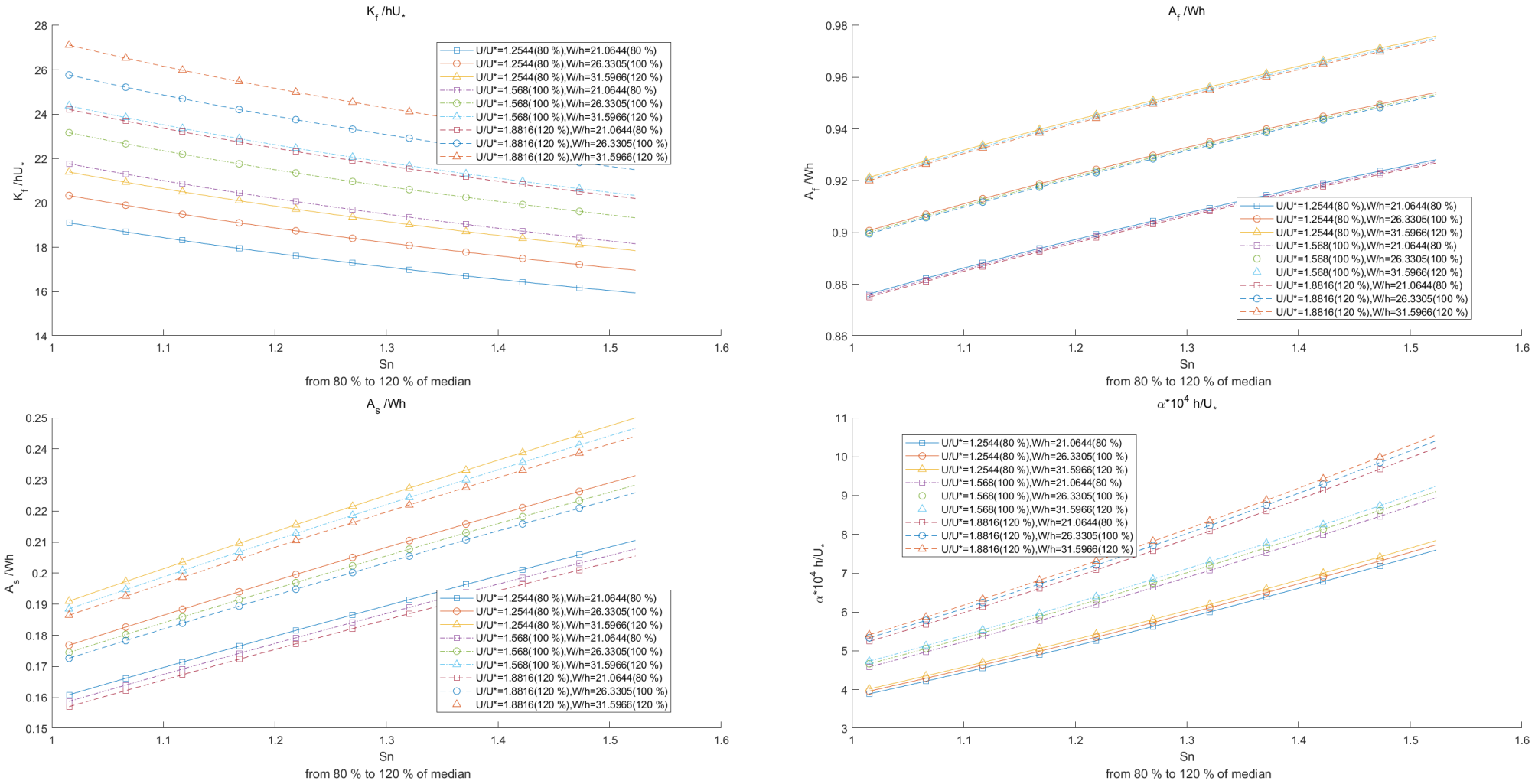


Figure 5.10 Calculated parameter values in a range of 20 % around the median of S_n (PCR)

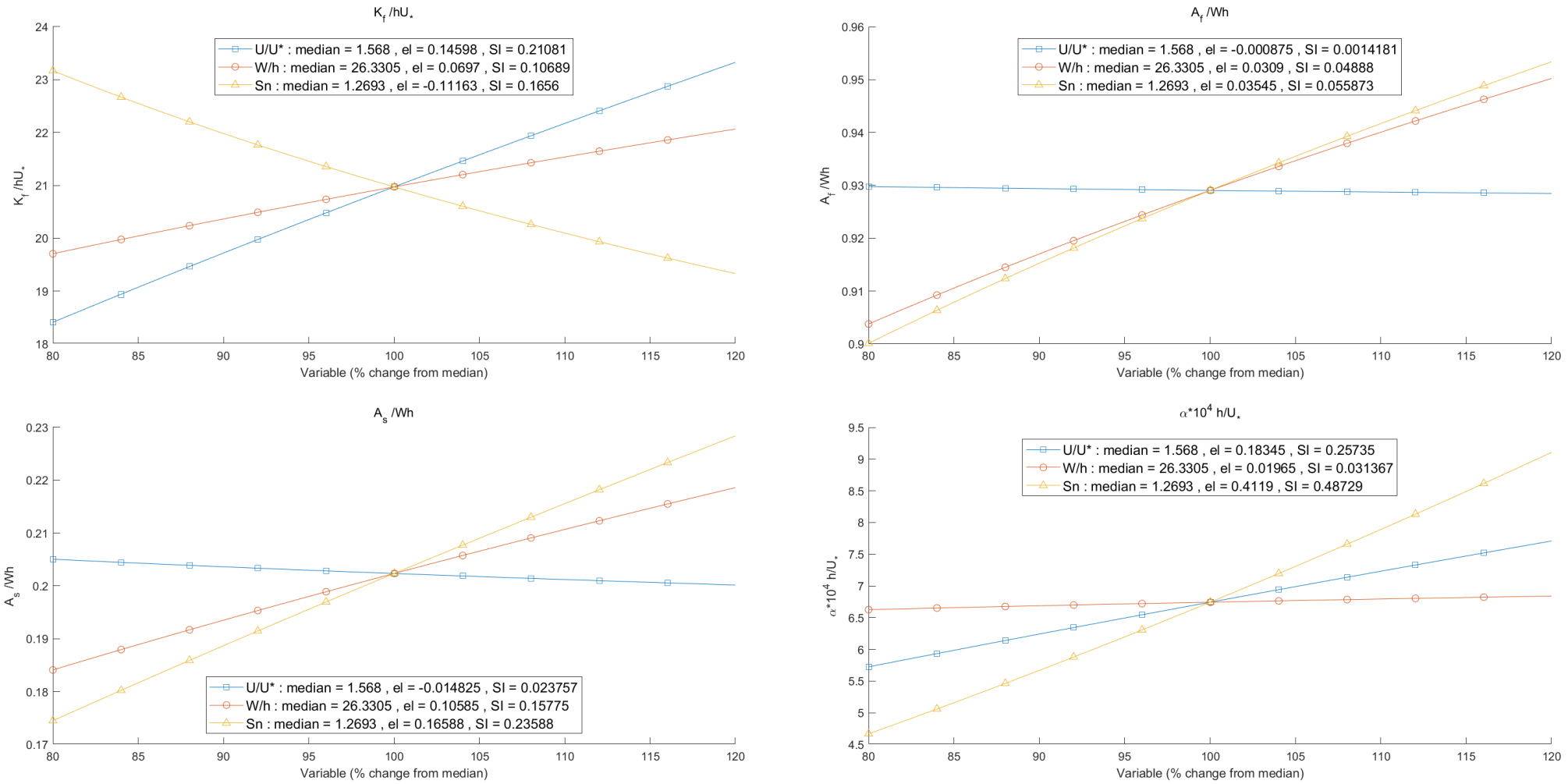


Figure 5.11 Spider graph of each TSM parameter and calculated sensitivity criterions (PCR)

5.3 In-stream Application of Empirical Equations

Using concentration data from the tracer test conducted on Cheong-Mi Creek, 6 sets of TSM parameters were obtained (Table 5.2). In order to confirm the applicability of developed TSM parameters, several solute transport simulation was produced using predicted TSM parameter using empirical equations. Simulated BTCs are depicted in Figure 5.12 - 17. The hydromorphic features of benchmarked experiment case are identical to Figure 3.4 and Table 3.1. Since free flow zone area A_f , accounting for flow velocity, is assumed as a known parameter in Femeena et al. (2019)'s model, the A_f values were calculated by Q/U where the A_f is not determined by the empirical equation.

According to left side plots of Figures 5.12 - 17, Femeena et al. (2019) model's prediction was much accurate than Eq. (4.5-8) except for reach 1-3 and 1-4. Eq. (4.9-12) produced intermediate shaped BTCs. Between section 2 and 4, meandering bend with point bar and the Hyun-Sa Bridge are exist varying hydraulic properties. As shown in Figure 5.8, since mean flow velocity along reach is not linear, the mean flow velocity is not identical to averaged cross-sectional velocity. From left hand side plots, stream with complex flow distribution such as reach 1-3 and 1-4, determining the free flow zone area A_f using Eq. (4.6) can be helpful for a rough approximation.

In order to compare BTC predictability of Eq. (4.5-8), Eq. (4.9-12) and Eq. (2.48-50) under the same condition, further simulations were conducted taking A_f as averaged value in right hand side of Figure 5.9, in addition to calculated value using Eq. (4.6). By interpreting right hand side plots, in-stream applicability taking only mixing processes into account can be assessed. Where the advection is same, most prediction results of the developed model in this

study were much similar to measured BTCs than Femeena et al. (2019)'s model except for reach 2-4. The reason for low performance in reach 2-4 is over-estimation of channel complexity. Especially, the PCR model produced best BTCs where same A_f was applied. From the comparisons, where the A_f is taken as mean cross-sectional velocity, the model developed in this study showed the better result in describing a modification of BTC caused by the storage zone effect. However, if the A_f is predicted at the same time,

As shown in Table 5.2, predicted K_f values using Eq. (2.48-50) are overestimated than calibrated values while values of the storage zone area A_s and exchange ratio α were underestimated likewise discussed above section 5.1. Contrary to Eq. (2.48-50), the MGGP model overestimated K_f . As keep shown, the PCR model predicts intermediate value between other two models. As a result of intermediate predictability, calculated values using the PCR model were comparatively similar to the calibrated parameter values than other models.

According to supplementary data of Femeena et al. (2019)'s paper for the derivation of Eq. (2.48-50), derivation datasets of Femeena et al. (2019) include parameter estimation results of other solute transport model's than OTIS based parameter values. Especially, they used the dispersion coefficient of 1D-ADE, K . Cheong et al. (2003) remarked that the dispersion coefficient of 1D-ADE, K , is not identical to the K_f of TSM because storage zone exchange process is contained in the K while the TSM considers the storage zone exchange process and free flow zone dispersion effect separately. In this regard, overestimation of the K_f value is due to the adoption of 1D-ADE's dispersion coefficient in derivation. In the perspective of A_s and α , underestimation of these storage parameters may come from the exclusion of the

friction factors, sinuosity, expansion, and contradiction of channel, and etc. affecting STS and HTS, as summarized by Jackson et al. (2013).

Misprediction of the free flow zone area A_f is subjected to a fundamental weakness of TSM. Due to the lumped model characteristic, hydrologic structures along stream are simplified as free flow zone area. Although calculating solute transport simulation based on hydraulic condition calculated from hydro-dynamic model gives an accurate result, predictability of mean velocity was observed from in-stream application in this section. This means velocity should be spatiotemporally determined based on fair approximation. For example, if a user has a plan view of experiment site, users can approximate spatial velocity distribution under the assumption of that cross-section area is proportional to width measure from the plan view. The site specificity of TSM parameters is on the fact that the TSM is the lumped model which is every advection and mixing characteristics as a set of TSM parameter for a target reach. If at least velocity field is approximated in fair reason this site specificity would be diminished.

Table 5.2 Estimated TSM parameter results in Cheong-mi Creek

Reach	Method	TSM parameters				DaI
		K_f (m^2/s)	$A_f\ (m^2)$	$A_s\ (m^2)$	$\alpha * 10^4\ (1/s)$	
1-2	Cal.	1.3224	9.6304	5.4102	2.4118	2.143
	MGGP.	0.5628	17.1558	2.7490	2.0688	
	RPCR	1.1095	11.5430	2.5608	1.1601	
	Fem.	6.3913	**9.0400***	1.0629	0.2117	
1-3	Cal.	1.7023	9.3698	3.7284	1.5115	4.711
	MGGP.	0.6931	6.4688	1.6993	4.4975	
	RPCR	1.1828	5.5906	1.2963	3.2573	
	Fem.	5.6880	**5.6500***	1.1868	0.7032	
1-4	Cal.	1.9227	9.0703	3.2681	1.4195	5.614
	MGGP.	0.5666	8.4832	2.7679	3.5144	
	RPCR	1.0092	7.3695	1.8863	2.8577	
	Fem	5.9774	**6.9538***	1.3296	0.4513	
2-3	Cal.	1.2028	9.0363	2.9521	1.244	2.808
	MGGP.	0.5770	16.5184	2.5437	2.0550	
	RPCR	1.1641	11.2472	2.4077	1.1354	
	Fem	6.3780	**8.6923***	0.9839	0.2226	
2-4	Cal.	1.6357	8.7069	2.5015	1.1652	3.741
	MGGP.	0.4010	19.6129	3.7058	1.4803	
	RPCR	0.8525	13.7515	3.1811	0.8711	
	Fem.	5.4618	**12.2162***	1.0977	0.1336	
3-4	Cal.	2.1215	7.512	1.5307	1.5476	1.428
	MGGP.	0.5769	9.3218	1.9357	3.5597	
	RPCR	1.1281	7.0864	1.6199	2.1907	
	Fem	5.9005	**6.7463***	1.2024	0.4659	

*: calculated as $A_f = Q/U$

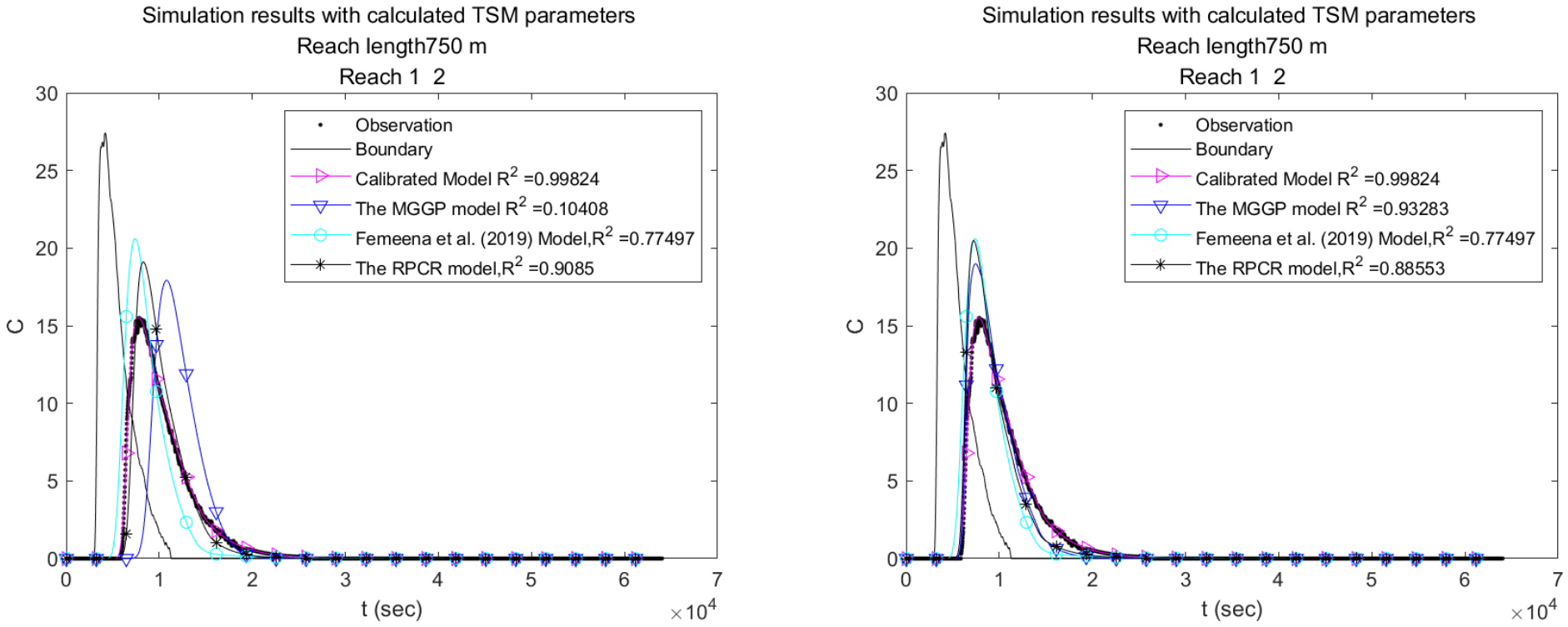
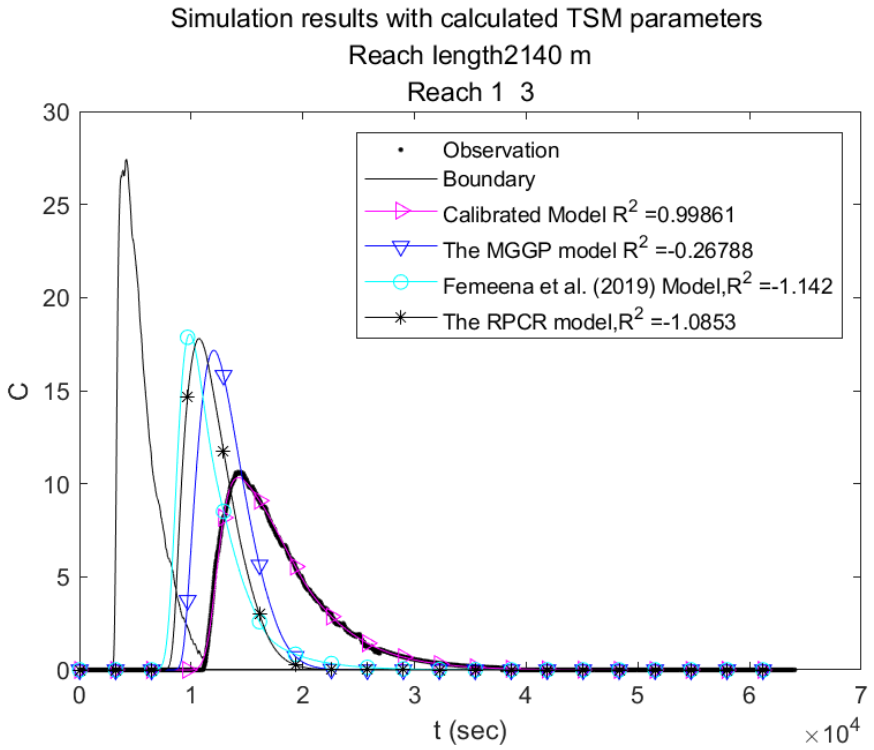


Figure 5.12 Plots of predicted BTCs using empirical equations (1-2)

(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6))



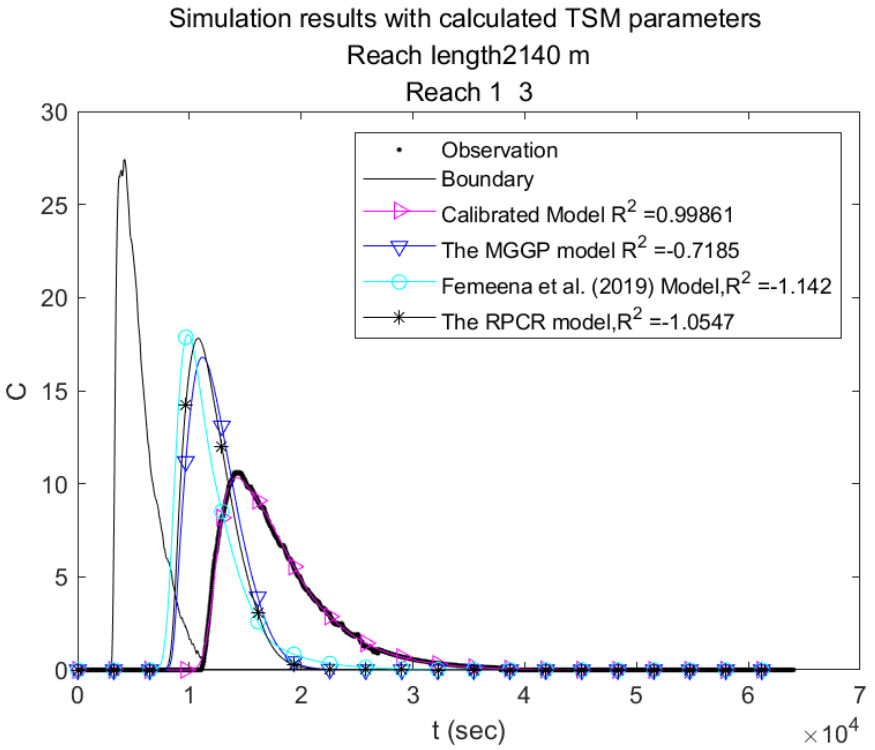


Figure 5.13 Plots of predicted BTCs using empirical equations (1-3)
(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6) and Eq(4.10))
(Continued)

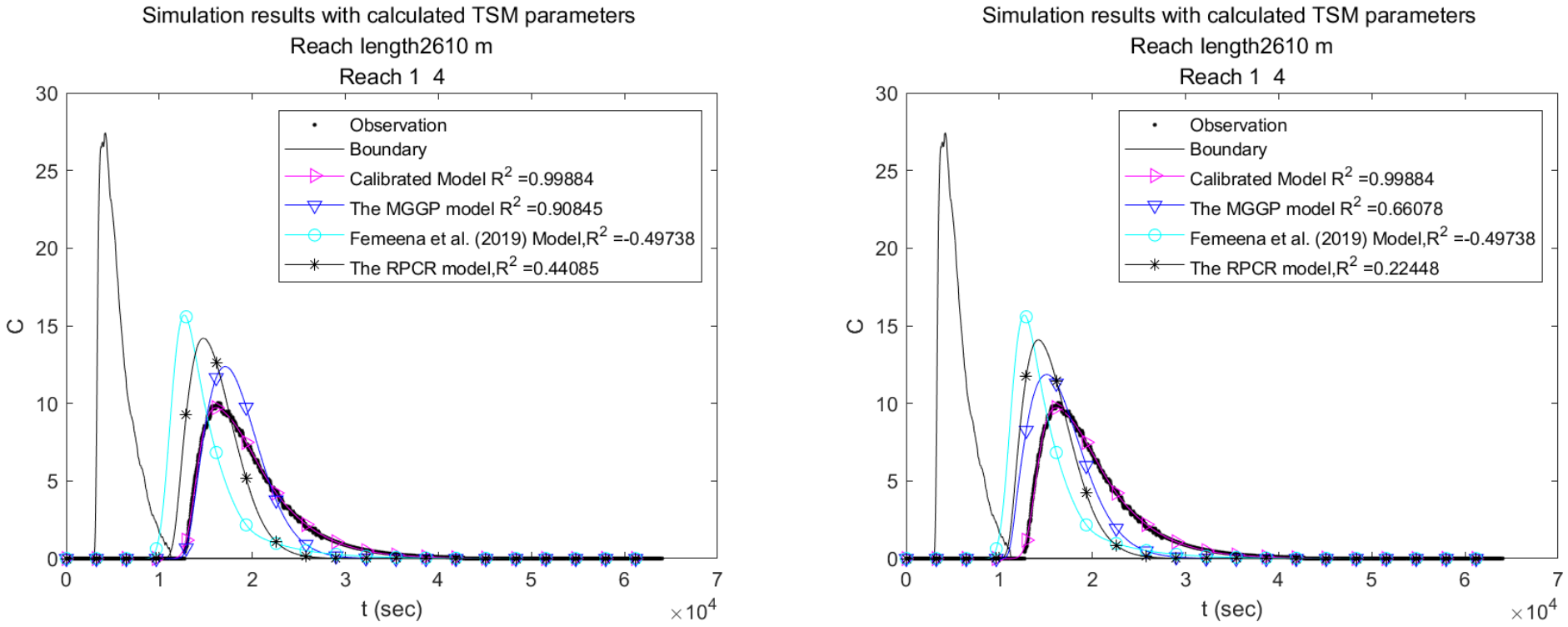
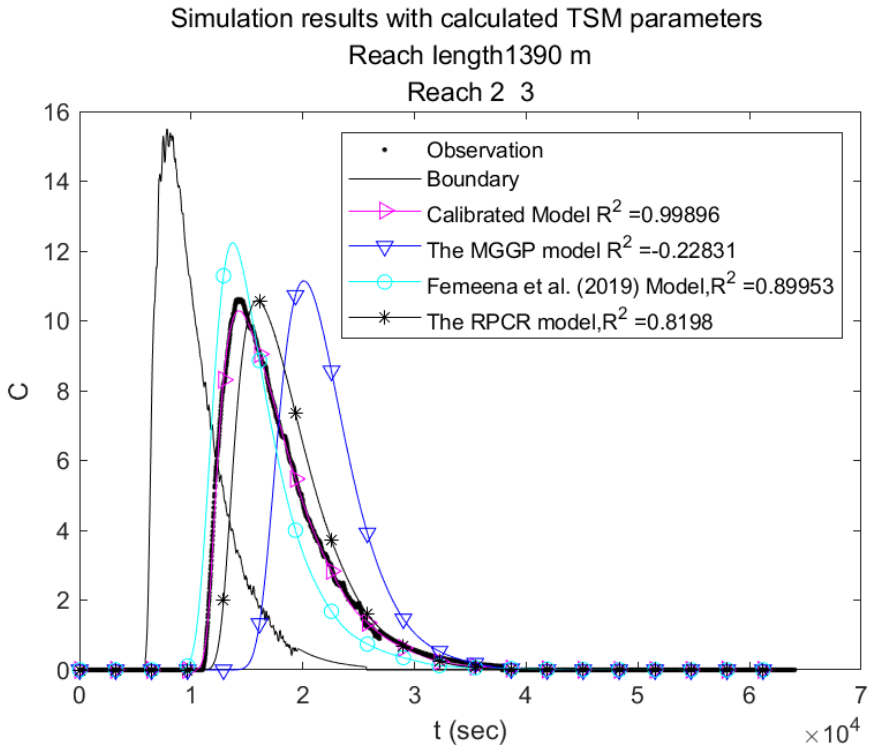


Figure 5.14 Plots of predicted BTCs using empirical equations (1-4)

(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6) and Eq(4.10))

(Continued)



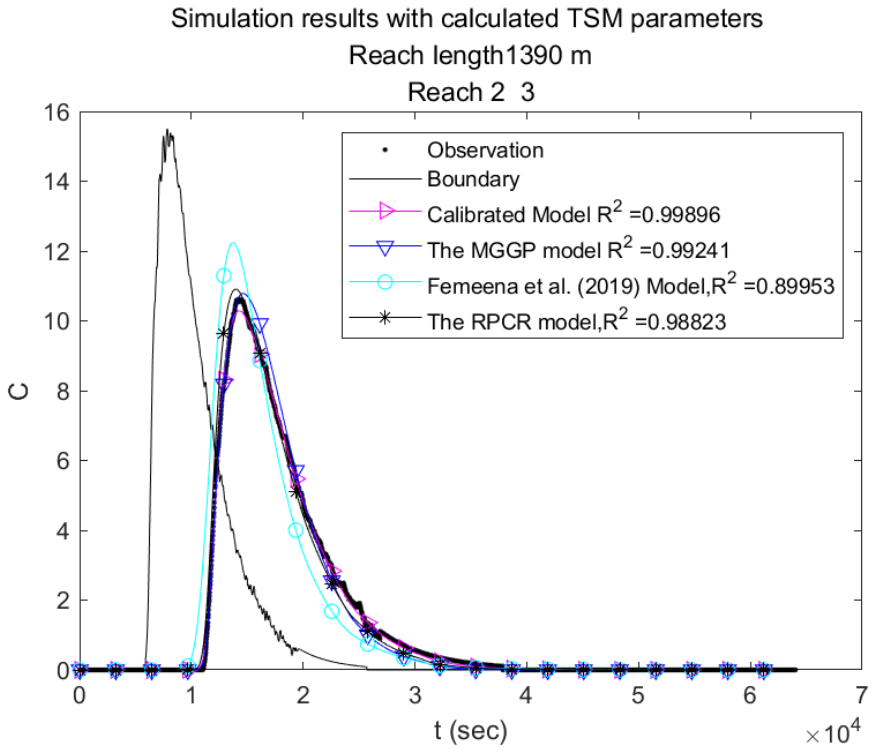


Figure 5.15 Plots of predicted BTCs using empirical equations (2-3)

(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6) and Eq(4.10))

(Continued)

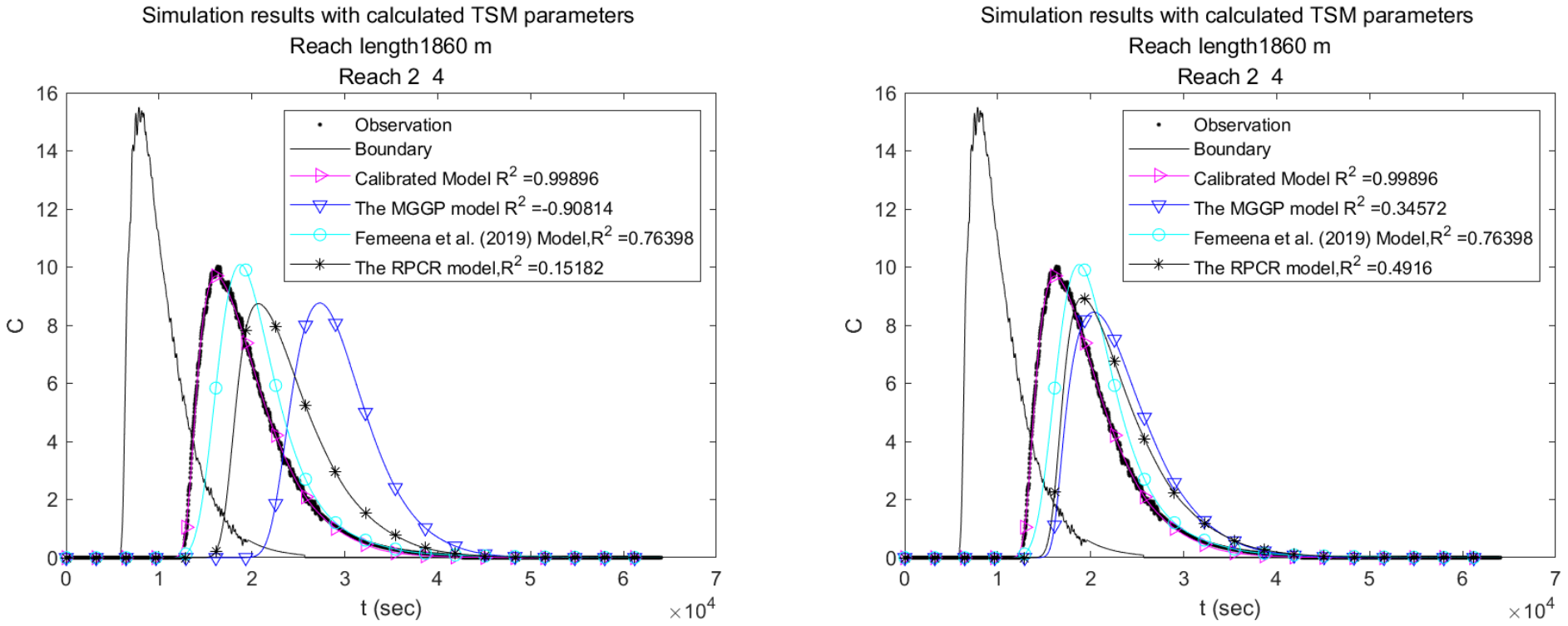
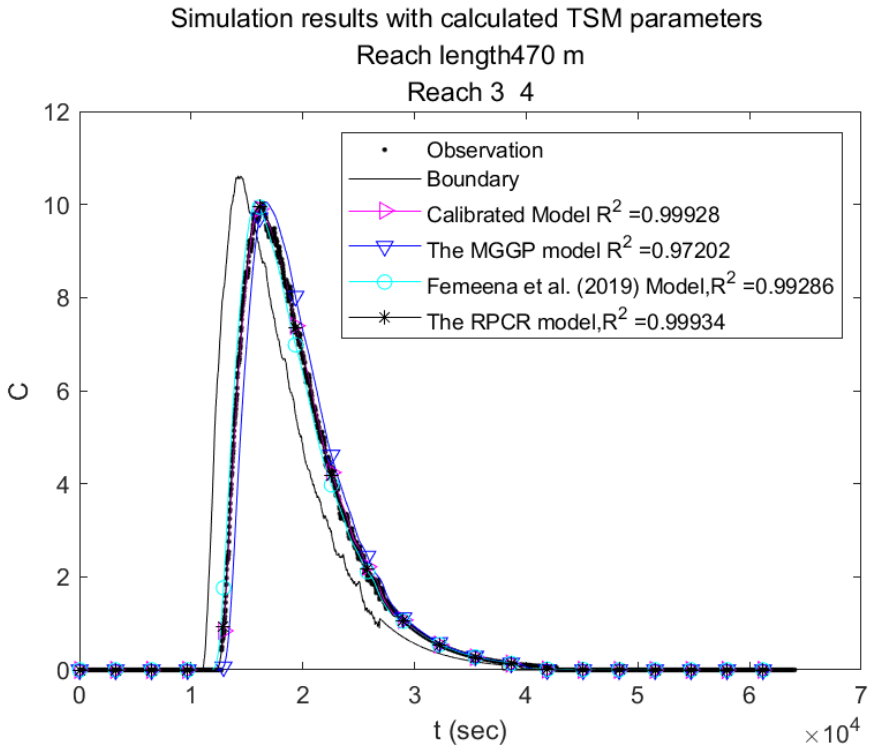


Figure 5.16 Plots of predicted BTCs using empirical equations (2-4)

(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6) and Eq(4.10))

(Continued)



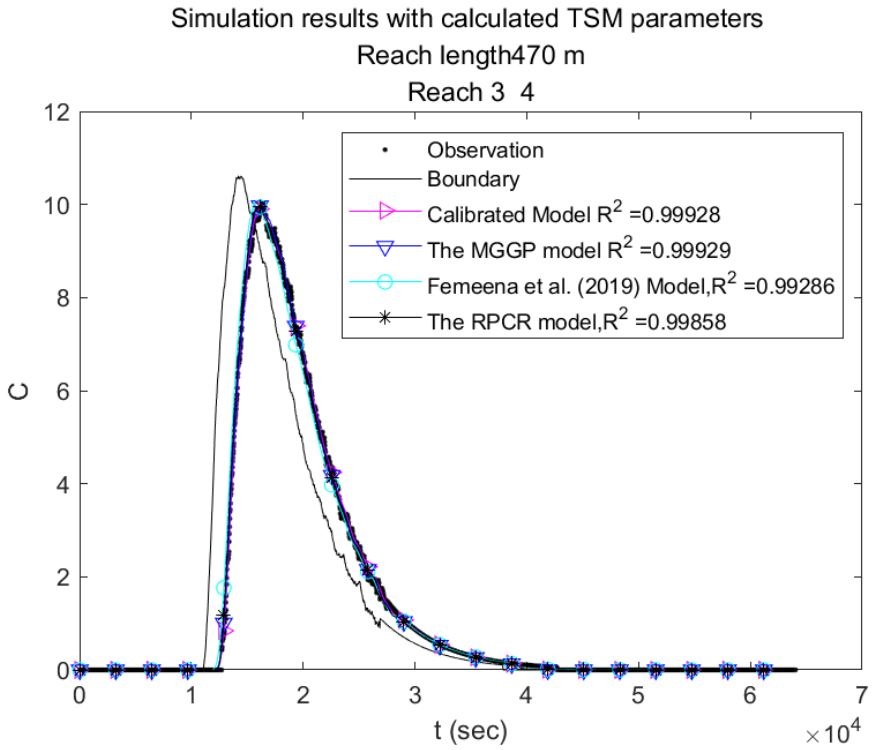


Figure 5.17 Plots of predicted BTCs using empirical equations (3-4)

(Left hand side $A_f = Q/U$; Right hand side: A_f = Eq(4.6) and Eq(4.10))

(Continued)

Chapter 6. Conclusion

In the present study, mainly two TSM parameter estimation methodologies were presented. First, the TSM parameter estimation framework was made by incorporating the TSM and SC-SAHEL optimization algorithm. In order to find the optimal condition of parameter estimation, several numerical experiments were conducted on Cheong-mi Creek tracer injection test result of NIER (2015). From numerical experiments, SC-SAHEL with a combination of EAs with MSE objective function was the most robust and efficient setting. Second, novel empirical equations for determining TSM parameters were proposed using both symbolic regression method MGGP and classical regression method PCR. In order to construct TSM parameter dataset, both meta-analysis and parameter estimation using TSM-SC-SAHEL were conducted. In order to develop generalized formulae, dimensionless TSM parameters and input variables were determined via dimensional analysis. Correlations between nominated dimensionless variables and TSM parameters were interpreted physically. Accordingly, eight empirical equations for TSM key parameters were derived via MGGP and PCR.

In order to suggest optimal parameter estimation framework, convergence assessment of optimization settings was carried out with respect to EAs and objective functions. Average parameter estimation success ratio values of the DE and the SC-SAHEL were 93 % and 88 %, these were first and second robust method, respectively. In addition to success ratio, EEN representing efficiency of optimization method was calculated. Calculated EEN value of the multi-EA SC-SAHEL was 52, which is superior to every method. In the objective function point of view, PMSE was the best fitness function in convergence and speed because success ratio and EEM were 99 % and 58, respectively. As a result of assessment, multi-EA SC-SAHEL with PMSE objective function comparing simulated BTC with measured BTS was suggested

as a TSM parameter estimation framework since the optimization setting was the best combination for TSM parameter estimation.

Performances of the presented empirical equation and previously developed equations were compared using collected 135 parameter sets used in training and testing of MGGP and PCR regression. According to performance comparison, equations developed in this study using MGGP were superior to other models in terms of accuracy except for the storage zone area A_s. In the aspect of R^2, calculated R^2 value of the MGGP for K_f was 0.692 in testing set, that was the best model among assess models. However, calculated R^2 value of the formula for prediction of the α in testing set, which is lower than zero, was lower than the published model of Femeena et al. (2019), which is 0.015. From plotted DR histogram, overestimation and underestimation of Femeena et al. (2019)'s model and the MGGP, respectively, for the K_f were observed. In terms of every performance criteria, PCR equations for TSM parameters showed intermediate performance between the MGGP model and Femeena et al. (2019)'s model in every performance criteria. Moreover, the PCR model kept showing stable performances that are close to best model for every TSM parameter.

Moreover, parameter sensitivity in TSM parameters was assessed by plotting estimated value around the median input values of the collected hydromorphic dataset. The OAT sensitivity analysis was performed by plotting calculated TSM parameters in order to understand the transient storage mass exchange process. From the analysis, relationships between TSM parameters and each input variables were interpreted by taking mass transport into mechanism account. The K_f and U/U_* was significantly correlated. While the MGGP model for K_f is only related with U/U_*, the PCR model for K_f is proportional to W/h and S_n and

negatively correlated with U/U_*. The S_n has positive relationships with the other three TSM parameters in two developed models in this study. The key to interpretation about the S_n was secondary flow in meandering channel. Moreover, the S_n and the W/h were important factors for three TSM parameters (A_f, A_s, and α), because these variables are accounting for the existence of STS and HTS.

Finally, then predicted BTCs using empirical equations and observed BTCs were compared in order to asses in-stream applicability of developed equations. In comparison, predicted curves using the developed model in this study (Eq. 4.5-8 and 4.9-12), are much similar to observed curves than another model where flow distribution is complicated. Especially, the PCR model consistently produced intermediated TSM parameter values and produced best BTCs in some cases. From the plotted result, probability of BTC prediction using empirical equations without parameter estimation was demonstrated. However, in the simple stream, the accuracy of prediction was lower than varying hydraulic condition. Especially, this trend was significant comparing BTCs under different flow conditions by calculating A_f via empirical equations and taking averaged value, respectively. In test reach with complex flow structure, cross-sectional averaged velocity could not predict advection accurately. Because approximating velocity distribution using two cross-sections averaged hydrologic features is difficult. Solving the hydro-dynamic model before using TSM, for reduction of uncertainty approximating A_f value, is recommendable. Even though inaccuracy of it is an ideal condition that A_f is solved based on the one-dimensional flow model. Thereafter conducting calibration.

The purpose of empirical equation development for TSM parameters is to propose the substitution method of direct parameter estimation in the situation of performing tracer

injection test is difficult, not for displacing of conventional methods. Taking situation for application as an example, in catchment-scale of study or in urban area study, conducting extensive tracer injection experiment is problematic, but TSM parameters can be determined using proposed empirical equations. In parameter estimation point of view, using developed TSM empirical equations, initial values or parameter space range of optimization for parameter estimation can be approximated. Furthermore, suggested TSM parameter estimation framework TSM-SC-SAHEL will be helpful for overcoming local minimum problem of OTIS. On the other hand, difficulty for producing BTCs by using TSM empirical equations was observed in this study especially around artificial obstacles that disturbing natural river flow. Moreover, existence of artificial transient storages make analysis much difficult when a researcher wants to discern effects of certain transient storage over a reach. For that reason, more investigation of transient storage around construction, such as weir and pier, is necessary to water resource management in river system.

References

Alizadeh, M. J., Ahmadyar, D., & Afghantoloee, A. (2017). Improvement on the existing equations for predicting longitudinal dispersion coefficient. *Water resources management*, *31*(6), 1777-1794.

Beer, T., & Young, P. C. (1983). Longitudinal dispersion in natural streams. *Journal of environmental engineering*, *109*(5), 1049-1067.

Bencala, K. E., & Walters, R. A. (1983). Simulation of solute transport in a mountain pool-and-riffle stream: A transient storage model. *Water Resources Research*, *19*(3), 718-724.

Bencala, K. E., Gooseff, M. N., & Kimball, B. A. (2011). Rethinking hyporheic flow and transient storage to advance understanding of stream-catchment connections. *Water Resources Research*, *47*(3).

Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression diagnostics. J.

Beltaos, S., & Day, T. J. (1978). A field study of longitudinal dispersion. *Canadian Journal of Civil Engineering*, *5*(4), 572-585.

Berkowitz, B., & Scher, H. (1995). On characterization of anomalous dispersion in porous and fractured media. *Water Resources Research*, *31*(6), 1461-1466.

Boano, F., Revelli, R., & Ridolfi, L. (2005). Source identification in river pollution problems: A geostatistical approach. *Water resources research*, *41*(7).

Boano, F., Packman, A. I., Cortis, A., Revelli, R., & Ridolfi, L. (2007). A continuous time random walk approach to the stream transport of solutes. *Water Resources Research*, *43*(10).

Boano, F., Harvey, J. W., Marion, A., Packman, A. I., Revelli, R., Ridolfi, L., & Wörman, A. (2014). Hyporheic flow and transport processes: Mechanisms, models, and biogeochemical

implications. *Reviews of Geophysics*, *52*(4), 603-679.

Bohrman, K. J., & Strauss, E. A. (2018). Macrophyte-driven transient storage and phosphorus uptake in a western Wisconsin stream. *Hydrological Processes*, *32*(2), 253-263.

Bottacin-Busolin, A., Singer, G., Zaramella, M., Battin, T. J., & Marion, A. (2009). Effects of streambed morphology and biofilm growth on the transient storage of solutes. *Environmental science & technology*, *43*(19), 7337-7342.

Boulton, A. J., Findlay, S., Marmonier, P., Stanley, E. H., & Valett, H. M. (1998). The functional significance of the hyporheic zone in streams and rivers. *Annual Review of Ecology and Systematics*, *29*(1), 59-81.

Briggs, M. A., Gooseff, M. N., Arp, C. D., & Baker, M. A. (2009). A method for estimating surface transient storage parameters for streams with concurrent hyporheic storage. *Water Resources Research*, *45*(4).

Brooks, B. W., Lazorchak, J. M., Howard, M. D., Johnson, M. V. V., Morton, S. L., Perkins, D. A., ... & Steevens, J. A. (2016). Are harmful algal blooms becoming the greatest inland water quality threat to public health and aquatic ecosystems?. *Environmental toxicology and chemistry*, *35*(1), 6-13.

Buffington, J. M., & Tonina, D. (2009). Hyporheic exchange in mountain rivers II: effects of channel morphology on mechanics, scales, and rates of exchange. *Geography Compass*, *3*(3), 1038-1062.

Carrera, J., Sánchez-Vila, X., Benet, I., Medina, A., Galarza, G., & Guimerà, J. (1998). On matrix diffusion: formulations, solution methods and qualitative effects. *Hydrogeology Journal*, *6*(1), 178-190.

Chatwin, P. C. (1980). Presentation of longitudinal dispersion data. *Journal of the Hydraulics Division*, *106*(1), 71-83.

Cheong, T. S., & Seo, I. W. (2003). Parameter estimation of the transient storage model by a routing method for river mixing processes. *Water resources research*, *39*(4).

Cheong, T. S., Younis, B. A., & Seo, I. W. (2007). Estimation of key parameters in model for solute transport in rivers and streams. *Water resources management*, *21*(7), 1165-1186.

Choi, J., Harvey, J. W., & Conklin, M. H. (2000). Characterizing multiple timescales of stream and storage zone interaction that affect solute fate and transport in streams. *Water Resources Research*, *36*(6), 1511-1518.

Chu, W., Gao, X., & Sorooshian, S. (2011). A new evolutionary search strategy for global optimization of high-dimensional problems. *Information Sciences*, *181*(22), 4909-4927.

Cohon, J. L., & Marks, D. H. (1975). A review and evaluation of multiobjective programing techniques. *Water Resources Research*, *11*(2), 208-220.

Czernuszenko, W., RowiŃSki, P. M., & Sukhodolov, A. (1998). Experimental and numerical validation of the dead-zone model for longitudinal dispersion in rivers. *Journal of Hydraulic Research*, *36*(2), 269-280.

Davis, P. M., & Atkinson, T. C. (2000). Longitudinal dispersion in natural channels: 3. An aggregated dead zone model applied to the River Severn, UK. *Hydrol Earth Syst SC*, *4*(3), 373-381.

Davidson, J. W., Savic, D., & Walters, G. A. (1999). Method for the identification of explicit polynomial formulae for the friction in turbulent pipe flow. *Journal of Hydroinformatics*, *1*(2), 115-126.

Davidson, J. W., Savic, D. A., & Walters, G. A. (2000). Approximators for the Colebrook-White formula obtained through a hybrid regression method. Computational Methods in Water Resources, Bentley.

Day, T. J. (1975). Longitudinal dispersion in natural channels. *Water Resources Research*, *11*(6), 909-918.

Deng, Z. Q., Singh, V. P., & Bengtsson, L. (2004). Numerical solution of fractional advection-dispersion equation. *Journal of Hydraulic Engineering*, *130*(5), 422-431.

Disley, T., Gharabaghi, B., Mahboubi, A. A., & McBean, E. A. (2015). Predictive equation for longitudinal dispersion coefficient. *Hydrological processes*, *29*(2), 161-172.

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of optimization theory and applications*, *76*(3), 501-521.

Dybas, C. L. (2005). Dead zones spreading in world oceans. *AIBS Bulletin*, *55*(7), 552-557.

Elliott, A. H., & Brooks, N. H. (1997). Transfer of nonsorbing solutes to a streambed with bed forms: Theory. *Water Resources Research*, *33*(1), 123-136.

Ensign, S. H., & Doyle, M. W. (2005). In-channel transient storage and associated nutrient retention: Evidence from experimental manipulations. *Limnology and Oceanography*, *50*(6), 1740-1751.

Etemad-Shahidi, A., & Taghipour, M. (2012). Predicting longitudinal dispersion coefficient in natural streams using M5′ model tree. *Journal of hydraulic engineering*, *138*(6), 542-554.

Eusuff, M. M., & Lansey, K. E. (2003). Optimization of water distribution network design using the shuffled frog leaping algorithm. *Journal of Water Resources planning and*

management, *129*(3), 210-225.

Eusuff, M., Lansey, K., & Pasha, F. (2006). Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization. *Engineering optimization*, *38*(2), 129-154.

Femeena, P. V., Chaubey, I., Aubeneau, A., McMillan, S., Wagner, P. D., & Fohrer, N. (2019). Simple regression models can act as calibration-substitute to approximate transient storage parameters in streams. *Advances in Water Resources*, *123*, 201-209.

Fischer, H. B. (1966). Longitudinal dispersion in laboratory and natural streams.

Fischer, H. B., List, E. J., & Koh, R. C. Y., Imberger, J. and Brooks, N. H.(1979) Mixing in Inland and Coastal Waters.

Fischer, H. B., List, J. E., Koh, C. R., Imberger, J., & Brooks, N. H. (2013). *Mixing in inland and coastal waters*. Elsevier.

Gandomi, A. H., Alavi, A. H., Arjmandi, P., Aghaeifar, A., & Seyednour, R. (2010). Genetic programming and orthogonal least squares: a hybrid approach to modeling the compressive strength of CFRP-confined concrete cylinders. *Journal of Mechanics of Materials and Structures*, *5*(5), 735-753.

Gandomi, A. H., & Alavi, A. H. (2012). A new multi-gene genetic programming approach to nonlinear system modeling. Part I: materials and structural engineering problems. *Neural Computing and Applications*, *21*(1), 171-187.

Ghane, A., Mazaheri, M., & Samani, J. M. V. (2016). Location and release time identification of pollution point source in river networks based on the Backward Probability Method. *Journal of environmental management*, *180*, 164-171.

Giustolisi, O., Doglioni, A., Savic, D. A., & Laucelli, D. (2004). A proposal for an effective

multiobjective non-dominated genetic algorithm: the OPTimised Multi-Objective Genetic Algorithm, OPTIMOGA. *OPTIMOGA, Report*, *7*.

Giustolisi, O., & Savic, D. A. (2006). A symbolic data-driven technique based on evolutionary polynomial regression. *Journal of Hydroinformatics*, *8*(3), 207-222.

Giustolisi, O., & Savic, D. A. (2009). Advances in data-driven analyses and modelling using EPR-MOGA. *Journal of Hydroinformatics*, *11*(3-4), 225-236.

Goldberg, D. E. (1989). Genetic Alogorithms in Search. *Optimization & Machine Learning*.

Gooseff, M. N., Briggs, M. A., Bencala, K. E., McGlynn, B. L., & Scott, D. T. (2013). Do transient storage parameters directly scale in longer, combined stream reaches? Reach length dependence of transient storage interpretations. *Journal of hydrology*, *483*, 16-25.

Graf, J. B. (1995). MEASURED AND PREDICTED VELOCITY AND LONGITUDINAL DISPERSION AT STEAI) Y AND UNSTEADY FLOW, COLORADO RIVER, GLEN CANYON DAM TO LAKE MEAD 1. *JAWRA Journal of the American Water Resources Association*, *31*(2), 265-281.

Haggerty, R., McKenna, S. A., & Meigs, L. C. (2000). On the late-time behavior of tracer test breakthrough curves. *Water Resources Research*, *36*(12), 3467-3479.

Haimes, Y. Y., & Hall, W. A. (1977). Sensitivity, responsivity, stability and irreversibility as multiple objectives in civil systems. *Advances in Water Resources*, *1*(2), 71-81.

Hart, D. R. (1995). Parameter estimation and stochastic interpretation of the transient storage model for solute transport in streams. *Water Resources Research*, *31*(2), 323-328.

Harvey, J. W., Wagner, B. J., & Bencala, K. E. (1996). Evaluating the reliability of the stream tracer approach to characterize stream-subsurface water exchange. *Water resources research*,

32(8), 2441-2451.

Harvey, J. W. (2000). Quantifying hydrologic interactions between streams and their subsurface hyporheic structure. *Streams and ground waters*, 3-44.

Harvey, J. W., Saiers, J. E., & Newlin, J. T. (2005). Solute transport and storage mechanisms in wetlands of the Everglades, south Florida. *Water Resources Research, 41*(5).

Hays, J. R. (1967). MASS TRANSPORT MECHANISMS IN OPEN CHANNEL FLOW.

Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: a new approach to robust principal component analysis. *Technometrics*, *47*(1), 64-79.

Jackson, T. R., Haggerty, R., & Apte, S. V. (2013). A fluid-mechanics based classification scheme for surface transient storage in riverine environments: quantitatively separating surface from hyporheic transient storage. *Hydrology and Earth System Sciences*, *17*(7), 2747-2779.

Kelleher, C., Wagener, T., McGlynn, B., Ward, A. S., Gooseff, M. N., & Payn, R. A. (2013). Identifiability of transient storage model parameters along a mountain stream. *Water Resources Research*, *49*(9), 5290-5306.

Kerr, P. C., Gooseff, M. N., & Bolster, D. (2013). The significance of model structure in one-dimensional stream solute transport models with multiple transient storage zones–competing vs. nested arrangements. *Journal of hydrology*, *497*, 133-144.

Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection* (Vol. 1). MIT press.

Koza, J. R. (1994). *Genetic programming II* (Vol. 17). Cambridge, MA: MIT press.

Laucelli, D., & Giustolisi, O. (2011). Scour depth modelling by a multi-objective evolutionary paradigm. *Environmental Modelling & Software*, *26*(4), 498-509.

Margolin, G., Dentz, M., & Berkowitz, B. (2003). Continuous time random walk and multirate mass transfer modeling of sorption. *Chemical physics*, *295*(1), 71-80.

Mariani, V. C., Luvizotto, L. G. J., Guerra, F. A., & dos Santos Coelho, L. (2011). A hybrid shuffled complex evolution approach based on differential evolution for unconstrained optimization. *Applied Mathematics and Computation*, *217*(12), 5822-5829.

Marion, A., & Zaramella, M. (2005). A residence time model for stream-subsurface exchange of contaminants. *Acta Geophysica Polonica*, *53*(4), 527.

Marion, A., Zaramella, M., & Bottacin-Busolin, A. (2008). Solute transport in rivers with multiple storage zones: The STIR model. *Water resources research*, *44*(10).

McQuivey, R. S., & Keefer, T. N. (1974). Simple method for predicting dispersion in streams. *Journal of the Environmental Engineering Division*, *100*(4), 997-1011.

Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46-61.

Montroll, E. W., & Weiss, G. H. (1965). Random walks on lattices. II. *Journal of Mathematical Physics*, *6*(2), 167-181.

Mulholland, P. J., Marzolf, E. R., Webster, J. R., Hart, D. R., & Hendricks, S. P. (1997). Evidence that hyporheic zones increase heterotrophic metabolism and phosphorus uptake in forest streams. *Limnology and oceanography*, *42*(3), 443-451.

Naeini, M. R., Yang, T., Sadegh, M., AghaKouchak, A., Hsu, K. L., Sorooshian, S., ... & Lei, X. (2018). Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL) optimization framework. *Environmental Modelling & Software*, *104*, 215-235.

Najafzadeh, M., Laucelli, D. B., & Zahiri, A. (2017). Application of model tree and

evolutionary polynomial regression for evaluation of sediment transport in pipes. *KSCE Journal of Civil Engineering*, *21*(5), 1956-1963.

Nagaoka, H., & Ohgaki, S. (1990). Mass transfer mechanism in a porous riverbed. *Water Research*, *24*(4), 417-425.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313.

NIER (2015), *Development and Application of Validation Technique for Advection-Diffusion of Dissolved Pollutants Using Tracer Experiments (II)*, Ministry of environment (in Korean)

Nordin, C. F., & Sabol, G. V. (1974). *Empirical data on longitudinal dispersion in rivers* (No. 74-20). US Geological Survey.

Nordin, C. F., & Troutman, B. M. (1980). Longitudinal dispersion in rivers: The persistence of skewness in observed data. *Water Resources Research*, *16*(1), 123-128.

Liu, H., & Cheng, A. H. (1980). Modified Fickian model for predicting dispersion. *Journal of the Hydraulics Division*, *106*(6), 1021-1040.

O'Connor, B. L., Hondzo, M., & Harvey, J. W. (2009). Predictive modeling of transient storage and nutrient uptake: Implications for stream restoration. *Journal of Hydraulic Engineering*, *136*(12), 1018-1032.

Pannell, D. J. (1997). Sensitivity analysis: strategies, methods, concepts, examples. *Agric econ*, *16*, 139-152.

Pedersen, F. B., Prediction of longitudinal dispersion in natural streams, *Ser. Pap. 14*, Inst. of Hydrodyn. and Hydraul. Eng., Univ. of Denmark, Lyngby, Denmark, 1977.

Qin, A. K., & Suganthan, P. N. (2005, September). Self-adaptive differential evolution

algorithm for numerical optimization. In *Evolutionary Computation, 2005. The 2005 IEEE Congress on*(Vol. 2, pp. 1785-1791). IEEE.

Rana, S. M., Scott, D. T., & Hester, E. T. (2017). Effects of in-stream structures and channel flow rate variation on transient storage. *Journal of hydrology*, *548*, 157-169.

Reed, P. M., Hadka, D., Herman, J. D., Kasprzyk, J. R., & Kollat, J. B. (2013). Evolutionary multiobjective optimization in water resources: The past, present, and future. *Advances in water resources*, *51*, 438-456.

Rezaie-Balf, M., & Kisi, O. (2017). New formulation for forecasting streamflow: evolutionary polynomial regression vs. extreme learning machine. *Hydrology Research*, nh2017283.

Rowiński, P. M., Piotrowski, A., & Napiórkowski, J. J. (2005). Are artificial neural network techniques relevant for the estimation of longitudinal dispersion coefficient in rivers?/Les techniques de réseaux de neurones artificiels sont-elles pertinentes pour estimer le coefficient de dispersion longitudinale en rivières?. Hydrological sciences journal, 50(1).

ROWIŃSKI, P. M., & Piotrowski, A. (2008). Estimation of parameters of the transient storage model by means of multi-layer perceptron neural networks/Estimation des paramètres du modèle de transport TSM au moyen de réseaux de neurones perceptrons multi-couches. *Hydrological sciences journal*, *53*(1), 165-178.

Runkel, R. L., & Chapra, S. C. (1993). An efficient numerical solution of the transient storage equations for solute transport in small streams. *Water Resources Research*, *29*(1), 211-215.

Runkel, R. L. (1998). One-dimensional transport with inflow and storage (OTIS): A solute transport model for streams and rivers.

Rutherford, J. C. (1994). *River mixing*. John Wiley & Son Ltd.

Sabol, G. V., & Nordin, C. F. (1978). Dispersion in rivers as related to storage zones. *Journal of the Hydraulics Division*, *104*(5), 695-708.

Sadegh, M., & Vrugt, J. A. (2014). Approximate bayesian computation using markov chain monte carlo simulation: Dream (abc). *Water Resources Research*, *50*(8), 6767-6787.

Sahay, R. R. (2011). Prediction of longitudinal dispersion coefficients in natural rivers using artificial neural network. Environmental Fluid Mechanics, 11(3), 247-261.

Sahay, R. R. (2012). Predicting transient storage model parameters of rivers by genetic algorithm. *Water resources management*, *26*(13), 3667-3685.

Scher, H., & Lax, M. (1973). Stochastic transport in a disordered solid. I. Theory. *Physical Review B*, *7*(10), 4491.

Searson, D. P., Leahy, D. E., & Willis, M. J. (2010, March). GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In *Proceedings of the International multiconference of engineers and computer scientists* (Vol. 1, pp. 77-80). Hong Kong: IMECS.

Searson, D. P. (2015). GPTIPS 2: an open-source software platform for symbolic data mining. In *Handbook of genetic programming applications* (pp. 551-573). Springer, Cham.

Seo, I. W., & Yu, D. Y. (2000). CHARACTERIZATION OF POOL-RIFFLE SEQUENCES IN SOLUTE TRANSPORT MODELING OF STREAMS. *Water Engineering Research*, *1*(3), 171-185.

Singh, S. K. (2003). Treatment of stagnant zones in riverine advection-dispersion. *Journal of Hydraulic Engineering*, *129*(6), 470-473.

Singh, S. K. (2008). Comparing three models for treatment of stagnant zones in riverine transport. *Journal of Irrigation and Drainage Engineering*, *134*(6), 853-856.

Smits, G., Kordon, A., Vladislavleva, K., Jordaan, E., & Kotanchek, M. (2006). Variable selection in industrial datasets using pareto genetic programming. In *Genetic Programming Theory and Practice III* (pp. 79-92). Springer, Boston, MA.

Socolofsky, S. A., & Jirka, G. H. (2005). Special topics in mixing and transport processes in the environment. *Engineering–Lectures. 5th Edition. Texas A&M University*, 1-93.

Stonedahl, S. H., Harvey, J. W., & Packman, A. I. (2013). Interactions between hyporheic flow produced by stream meanders, bars, and dunes. *Water Resources Research*, *49*(9), 5450-5461.

Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, *11*(4), 341-359.

Szeftel, P., Moore, R. D., & Weiler, M. (2011). Influence of distributed flow losses and gains on the estimation of transient storage parameters from stream tracer experiments. *Journal of hydrology*, *396*(3-4), 277-291.

Tayfur, G. (2006). Fuzzy, ANN, and regression models to predict longitudinal dispersion coefficient in natural streams. Hydrology Research, 37(2), 143-164.

Thackston, E. L., & Schnelle, K. B. (1970). Predicting effects of dead zones on stream mixing. *Journal of the Sanitary Engineering Division*, *96*(2), 319-331.

Thomas, S. A., Valett, H. M., Webster, J. R., & Mulholland, P. J. (2003). A regression approach to estimating reactive solute uptake in advective and transient storage zones of stream ecosystems. *Advances in Water Resources*, *26*(9), 965-976.

Tonina, D., & Buffington, J. M. (2009). Hyporheic exchange in mountain rivers I: Mechanics and environmental effects. *Geography Compass*, *3*(3), 1063-1086.

Toprak, Z. F., & Cigizoglu, H. K. (2008). Predicting longitudinal dispersion coefficient in

natural streams by artificial intelligence methods. Hydrological Processes: An International Journal, 22(20), 4106-4129.

Triska, F. J., Kennedy, V. C., Avanzino, R. J., Zellweger, G. W., & Bencala, K. E. (1989). Retention and transport of nutrients in a third-order stream in northwestern California: Hyporheic processes. *Ecology*, *70*(6), 1893-1905.

TT Le, A., Kasahara, T., & Vudhivanich, V. (2018). Seasonal Variation and Retention of Ammonium in Small Agricultural Streams in Central Thailand. *Environments*, *5*(7), 78.

Valett, H. M., Morrice, J. A., Dahm, C. N., & Campana, M. E. (1996). Parent lithology, surface–groundwater exchange, and nitrate retention in headwater streams. *Limnology and oceanography*, *41*(2), 333-345.

Vaux, W. G. (1962). Interchange of stream and intragravel water in a salmon spawning riffle. *US Fish and Wildlife Service Special Scientific Report, Fisheries*, *405*, 11.

Versteeg, H. K., & Malalasekera, W. (2007). *An introduction to computational fluid dynamics: the finite volume method*. Pearson Education.

Vladislavleva, E. J., Smits, G. F., & Den Hertog, D. (2008). Order of nonlinearity as a complexity measure for models generated by symbolic regression via pareto genetic programming. *IEEE Transactions on Evolutionary Computation*, *13*(2), 333-349.

Vrugt, J. A., & Robinson, B. A. (2007). Improved evolutionary optimization from genetically adaptive multimethod search. *Proceedings of the National Academy of Sciences*, *104*(3), 708-711.

Vrugt, J. A., Robinson, B. A., & Hyman, J. M. (2008). Self-adaptive multimethod search for global optimization in real-parameter spaces. *IEEE Transactions on Evolutionary Computation*,

13(2), 243-259.

Wagner, B. J., & Harvey, J. W. (1997). Experimental design for estimating parameters of rate-limited mass transfer: Analysis of stream tracer studies. *Water Resources Research*, *33*(7), 1731-1741.

Ward, A. S., Kelleher, C. A., Mason, S. J., Wagener, T., McIntyre, N., McGlynn, B., ... & Payn, R. A. (2017). A software tool to assess uncertainty in transient-storage model parameters using Monte Carlo simulations. *Freshwater Science*, *36*(1), 195-217.

Ward, A. S., Morgan, J. A., White, J. R., & Royer, T. V. (2018). Streambed restoration to remove fine sediment alters reach-scale transient storage in a low-gradient fifth-order river, Indiana, USA. *Hydrological Processes*.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, *1*(1), 67-82.

Wörman, A., Packman, A. I., Johansson, H., & Jonsson, K. (2002). Effect of flow-induced exchange in hyporheic zones on longitudinal transport of solutes in streams and rivers. *Water Resources Research*, *38*(1), 2-1.

Wörman, A., Packman, A. I., Marklund, L., Harvey, J. W., & Stone, S. H. (2007). Fractal topography and subsurface water flows from fluvial bedforms to the continental shield. *Geophysical Research Letters*, *34*(7).

Yotsukura, N., Fischer, H. B., & Sayre, W. W. (1970). *Measurement of mixing characteristics of the Missouri River between Sioux City, Iowa, and Plattsmouth, Nebraska* (No. 1899-G). USGPO,.

Zaramella, M., Marion, A., Lewandowski, J., & Nützmann, G. (2016). Assessment of transient

storage exchange and advection–dispersion mechanisms from concentration signatures along breakthrough curves. *Journal of Hydrology*, *538*, 794-801.

Zhang, S. P., & Xin, X. K. (2017). Pollutant source identification model for water pollution incidents in small straight rivers based on genetic algorithm. *Applied Water Science*, *7*(4), 1955-1963

Appendix. I. The mean, minimum, and maximum values of the model fitness value and number of evolution using the SC-SAHEL with single-EA and multi-EA (Bold: best performance method)

Reach	Criteria	Obj. Func.	SCE-UA (SC-CCE)		SP-UCI (SC-MCCE)		SC-MFL		SC-DEF		SC-GWO		SC-SAHEL (CCE+MCCE+FL+DEF+GWO)	
			mean	range	mean	range	mean	range	mean	range	mean	range	mean	range
S1-S2	Model fitness	PMSE	**0.37**	0.37~0.37	2.15	0.37~5.72	**0.37**	0.37~0.37	**0.37**	0.37~0.37	**0.37**	0.37~0.37	**0.37**	0.37~0.37
		PME	14.2	5.18~18.7	18.7	18.7~18.7	9.68	5.18~18.7	**5.18**	5.18~5.18	5.23	5.23~5.23	9.68	5.18~18.7
		1-R2	1.35	0.23~3.59	1.35	0.23~3.59	**0.23**	0.23~0.23	**0.23**	0.23~0.23	**0.23**	0.23~0.23	1.35	0.23~3.59
		PMAE	3.33	3.32~3.34	2.35	0.84~3.34	**0.84**	0.84~0.84	2.51	0.84~3.34	2.5	0.84~3.34	2.51	0.84~3.34
	#of evolution	PMSE	71	31~93	42	33~59	53	46~59	66	57~73	94	89~101	**34**	33~35
		PME	60	41~71	68	63~76	58	52~68	69	66~72	155	146~166	**57**	41~74
		1-R2	**45**	31~59	54	46~60	46	42~48	53	50~56	88	82~93	49	40~65
		PMAE	57	56~58	**51**	43~55	105	82~137	78	56~120	107	67~184	59	54~67
S1-S3	Model fitness	PMSE	**0.2**	0.2~0.2	**0.2**	0.2~0.2	**0.2**	0.2~0.2	**0.2**	0.2~0.2	**0.2**	0.2~0.2	**0.2**	0.2~0.2
		PME	**4.5**	4.5~4.5	**4.5**	4.5~4.5	**4.5**	4.5~4.5	**4.5**	4.5~4.5	**4.5**	4.5~4.5	**4.5**	4.5~4.5
		1-R2	**0.18**	0.18~0.18	**0.18**	0.18~0.18	**0.18**	0.18~0.18	**0.18**	0.18~0.18	**0.18**	0.18~0.18	**0.18**	0.18~0.18
		PMAE	4.32	4.32~4.32	4.32	4.32~4.32	4.32	4.32~4.32	3.15	0.82~4.32	**2.77**	0.82~4.24	3.15	0.82~4.32
	#of evolution	PMSE	36	34~39	**32**	30~34	51	51~52	58	57~58	120	114~124	40	39~40
		PME	56	51~63	53	47~65	60	57~61	76	73~78	194	179~214	**51**	43~60
		1-R2	40	35~48	**30**	30~31	52	49~54	58	56~59	122	116~129	40	39~41
		PMAE	57	56~58	**56**	55~57	45	44~46	69	45~116	141	59~207	72	55~106
S1-S4	Model fitness	PMSE	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14
		PME	**4.15**	4.15~4.15	**4.15**	4.15~4.15	**4.15**	4.15~4.15	**4.15**	4.15~4.15	**4.15**	4.15~4.15	**4.15**	4.15~4.15
		1-R2	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13
		PMAE	3.89	3.89~3.89	3.89	3.89~3.89	3.6	3.02~3.89	**1.79**	0.74~3.89	3.47	2.74~3.89	3.87	3.83~3.89
	#of evolution	PMSE	44	43~45	**35**	33~36	57	56~58	66	63~71	163	138~209	46	40~52
		PME	71	61~84	70	60~80	**62**	55~69	94	88~101	239	219~251	67	56~86
		1-R2	42	41~44	**38**	37~40	54	50~59	67	65~68	152	137~167	50	49~52

		PMAE	56	56~57	56	56~57	56	44~77	95	57~126	76	56~112	57	56~59
S2-S3	Model fitness	PMSE	**0.16**	0.16~0.16	**0.16**	0.16~0.16	**0.16**	0.16~0.16	**0.16**	0.16~0.16	**0.16**	0.16~0.16	**0.16**	0.16~0.16
		PME	**3.46**	3.46~3.46	**3.46**	3.46~3.46	**3.46**	3.46~3.46	**3.46**	3.46~3.46	**3.46**	3.46~3.46	**3.46**	3.46~3.46
		1-R2	**0.15**	0.15~0.15	**0.15**	0.15~0.15	**0.15**	0.15~0.15	**0.15**	0.15~0.15	**0.15**	0.15~0.15	**0.15**	0.15~0.15
		PMAE	3.61	3.61~3.61	3.52	3.34~3.61	**0.93**	0.93~0.93	**0.93**	0.93~0.93	2.72	0.93~3.61	**0.93**	0.93~0.93
	#of evolution	PMSE	33	32~34	**29**	28~30	45	43~46	51	50~51	103	96~113	36	35~37
		PME	54	49~61	**46**	43~49	58	56~59	97	95~101	208	189~235	54	51~60
		1-R2	33	33~34	**28**	28~29	44	43~45	53	52~53	102	96~106	36	34~38
		PMAE	57	56~57	**69**	56~95	98	75~110	82	78~84	86	57~142	78	73~83
S2-S4	Model fitness	PMSE	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14	**0.14**	0.14~0.14
		PME	**3.25**	3.25~3.25	**3.25**	3.25~3.25	**3.25**	3.25~3.25	**3.25**	3.25~3.25	**3.25**	3.25~3.25	**3.25**	3.25~3.25
		1-R2	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13	**0.13**	0.13~0.13
		PMAE	3.29	3.29~3.29	3.24	3.12~3.29	3.29	3.29~3.29	**0.9**	0.9~0.9	3.1	2.72~3.29	**0.9**	0.9~0.9
	#of evolution	PMSE	35	34~36	**31**	30~32	49	46~50	59	58~61	113	109~118	40	38~43
		PME	58	49~62	**48**	43~54	55	53~57	85	81~89	190	179~201	54	49~60
		1-R2	36	34~37	**32**	31~32	50	48~53	59	59~59	114	110~122	39	39~40
		PMAE	56	55~57	**57**	54~60	47	46~48	82	77~85	77	55~119	71	55~80
S3-S4	Model fitness	PMSE	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08
		PME	**3.35**	3.35~3.35	**3.35**	3.35~3.35	**3.35**	3.35~3.35	**3.35**	3.35~3.35	3.37	3.35~3.39	**3.35**	3.35~3.35
		1-R2	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08	**0.08**	0.08~0.08
		PMAE	0.8	0.8~0.8	0.79	0.77~0.8	**0.62**	0.62~0.62	**0.62**	0.62~0.62	**0.62**	0.62~0.62	**0.62**	0.62~0.62
	#of evolution	PMSE	37	35~41	**31**	30~33	45	43~46	61	60~62	115	101~133	45	37~60
		PME	67	63~70	**57**	54~61	66	58~71	111	110~111	247	234~271	72	69~76
		1-R2	33	33~33	**31**	31~32	47	46~49	61	59~63	100	94~107	38	38~38
		PMAE	58	57~58	**56**	55~57	78	73~83	79	77~83	144	138~155	73	58~95

Appendix. II. Used dimensionless datasets for development of empirical equations

Reference	Test_reach	$\frac{K_f}{hU_*}$	$\frac{A_f}{Wh}$	$\frac{A_s}{Wh}$	$\frac{\alpha}{U_*/h}$	$\frac{U}{U_*}$	$\frac{W}{h}$	S_n	Testset
T. T. Le et el. (2018)	Kasetsart.Univ, Thailand	32.203	0.951	0.371	7.442	1.501	21.064	1.000	
T. T. Le et el. (2018)	Kasetsart.Univ, Thailand	6.092	0.711	0.114	6.435	1.443	14.615	1.000	T
T. T. Le et el. (2018)	Kasetsart.Univ, Thailand	18.125	0.850	0.060	7.722	2.405	14.615	1.000	
Bohrman and Strauss (2018)	Spring Coulee Creek, WI	17.389	0.998	0.070	1.595	2.483	22.714	1.080	T
Bohrman and Strauss (2018)	Spring Coulee Creek, WI	26.025	0.590	0.075	5.454	2.559	18.542	1.160	T
Bohrman and Strauss (2018)	Spring Coulee Creek, WI	14.045	0.491	0.117	8.052	2.037	10.718	1.220	
Bohrman and Strauss (2018)	Spring Coulee Creek, WI	20.961	1.348	0.341	5.175	2.901	14.571	1.220	
Bohrman and Strauss (2018)	Spring Coulee Creek, WI	13.090	0.599	0.056	8.108	2.867	9.634	1.280	
Claessens et al. (2010a)	Baisman Run, MD	10.631	1.034	0.207	0.981	0.485	21.714	1.009	
Claessens et al. (2010a)	Baisman Run, MD	28.292	1.001	0.143	0.639	0.878	32.121	1.130	
Claessens et al. (2010a)	Baisman Run, MD	43.158	1.006	0.210	1.901	1.115	45.352	1.135	
Claessens et al. (2010b)	Baisman Run, MD	11.122	0.978	0.133	0.868	0.482	26.615	1.044	T
Claessens et al. (2010b)	Baisman Run, MD	28.589	1.023	0.034	0.994	0.545	25.395	1.041	
Claessens et al. (2010b)	Baisman Run, MD	5.986	0.998	0.074	1.274	0.377	15.179	1.117	
Claessens et al. (2010b)	Baisman Run, MD	17.221	1.001	0.178	1.947	0.622	25.476	1.001	
Claessens et al. (2010b)	Baisman Run, MD	11.777	0.996	0.119	2.221	0.441	11.397	1.002	T
Claessens et al. (2010b)	Baisman Run, MD	15.169	0.982	0.145	1.385	0.527	21.818	1.010	
Claessens et al. (2010b)	Baisman Run, MD	3.608	1.004	0.367	3.227	0.386	26.190	1.010	
Claessens et al. (2010b)	Baisman Run, MD	8.519	1.005	0.228	10.307	1.062	27.640	1.031	T
Claessens et al. (2010b)	Baisman Run, MD	22.466	0.987	0.234	4.967	1.387	38.169	1.060	
Claessens et al. (2010b)	Baisman Run, MD	1.129	0.997	0.221	1.773	0.175	8.306	1.669	
Claessens et al. (2010b)	Baisman Run, MD	21.953	0.973	0.130	4.150	1.567	45.156	1.669	T
Claessens et al. (2010b)	Baisman Run, MD	21.228	1.010	0.460	7.870	1.408	38.194	1.028	
Claessens et al. (2010b)	Baisman Run, MD	36.754	0.988	0.194	17.898	1.648	51.757	1.055	
Claessens et al. (2010b)	Baisman Run, MD	19.908	1.003	0.119	1.036	0.699	32.756	1.240	T
Claessens et al. (2010b)	Baisman Run, MD	20.404	1.019	0.193	4.835	1.568	32.903	1.022	
Claessens et al. (2010b)	Baisman Run, MD	6.846	1.005	0.162	3.042	0.487	24.379	1.130	
Claessens et al. (2010b)	Baisman Run, MD	3.615	0.988	0.417	1.672	0.261	19.706	1.044	T
Claessens et al. (2010b)	Baisman Run, MD	22.821	1.043	0.302	0.366	0.473	20.735	1.041	
Claessens et al. (2010b)	Baisman Run, MD	18.249	0.980	0.284	0.627	0.441	15.556	1.117	
Claessens et al. (2010b)	Baisman Run, MD	18.274	1.055	0.188	0.646	0.643	29.259	1.001	
Claessens et al. (2010b)	Baisman Run, MD	10.249	1.020	0.102	0.527	0.482	14.938	1.002	
Claessens et al. (2010b)	Baisman Run, MD	6.029	1.023	0.300	1.648	0.369	20.864	1.010	
Claessens et al. (2010b)	Baisman Run, MD	4.566	0.965	0.263	6.515	1.068	31.667	1.031	
Claessens et al. (2010b)	Baisman Run, MD	22.592	0.992	0.319	0.625	0.794	23.205	1.060	
Claessens et al. (2010b)	Baisman Run, MD	1.342	1.016	0.451	1.381	0.769	32.143	1.669	T
Claessens et al. (2010b)	Baisman Run, MD	9.573	0.988	0.222	3.678	1.279	57.391	1.669	T

Claessens et al. (2010b)	Baisman Run, MD	23.163	1.020	0.145	4.789	1.477	63.889	1.055	
Claessens et al. (2010b)	Baisman Run, MD	74.673	1.029	0.544	4.493	1.343	58.644	1.240	
Claessens et al. (2010b)	Baisman Run, MD	6.075	0.991	0.301	8.573	0.945	40.000	1.022	
Claessens et al. (2010b)	Baisman Run, MD	93.269	0.984	0.241	0.579	1.301	36.133	1.130	
Gücker et al. (2009)	Pristine streams	12.569	0.933	0.367	19.728	2.864	7.500	1.100	
Gücker et al. (2009)	Aricultural streams	51.880	0.801	0.095	38.609	2.506	7.049	1.300	
Gücker et al. (2009)	Aricultural streams	306.337	0.953	0.299	11.861	8.219	7.455	1.100	
Stofleth et al. (2008)	Goodwin Creek, MS	104.439	0.713	0.226	27.980	2.623	39.917	1.120	
Stofleth et al. (2008)	Goodwin Creek, MS	228.308	0.800	0.783	3.497	3.497	39.917	1.120	T
Stofleth et al. (2008)	Goodwin Creek, MS	9.715	1.635	0.452	17.487	0.583	39.917	1.120	
Stofleth et al. (2008)	Goodwin Creek, MS	94.724	0.713	0.278	24.482	1.749	39.917	1.120	T
Bukaveckas (2007)	Wilson Creek, KY	58.490	0.979	0.275	31.871	4.277	39.286	1.103	
Bukaveckas (2007)	Wilson Creek, KY	73.277	1.424	0.577	20.674	5.672	75.556	1.098	T
Harvey et al. (2003)	Pinal Creek, AZ	67.604	1.035	0.118	4.133	6.118	37.474	1.225	
Harvey et al. (2003)	Pinal Creek, AZ	97.229	1.039	0.192	14.154	10.157	56.176	1.076	
Harvey et al. (2003)	Pinal Creek, AZ	179.288	1.007	0.351	3.970	4.517	87.143	1.062	
Johnson et al. (2014)	Truckee River, NV-CA	27.151	0.952	0.232	7.494	1.790	42.879	1.388	
Johnson et al. (2014)	Truckee River, NV-CA	12.664	0.815	0.316	23.232	3.103	49.905	1.444	
Johnson et al. (2014)	Truckee River, NV-CA	74.955	0.998	0.216	8.848	3.041	41.322	1.123	
Johnson et al. (2014)	Truckee River, NV-CA	23.162	0.756	0.220	18.849	1.658	43.391	1.220	
Muller Price et al. (2016)	Sheep Creek, CO	9.256	0.609	0.219	31.170	1.671	28.174	1.110	
Muller Price et al. (2016)	Sheep Creek, CO	60.395	1.111	0.478	0.343	1.659	40.907	1.240	
Muller Price et al. (2016)	Sheep Creek, CO	13.719	0.811	0.195	9.290	0.877	39.027	1.240	
Muller Price et al. (2016)	Sheep Creek, CO	14.004	1.089	0.305	23.300	1.608	24.200	1.630	T
Muller Price et al. (2016)	Sheep Creek, CO	7.478	0.915	0.156	14.461	0.725	15.254	1.490	
Muller Price et al. (2016)	Sheep Creek, CO	23.726	0.946	0.435	26.098	1.558	29.432	1.630	
Muller Price et al. (2016)	Nunn Creek, CO	5.078	0.991	0.198	19.405	1.021	22.727	1.610	
Muller Price et al. (2016)	Nunn Creek, CO	17.775	0.769	0.231	24.068	1.500	35.558	1.440	
Muller Price et al. (2016)	Nunn Creek, CO	6.927	0.942	0.198	14.426	0.697	20.346	1.610	
Muller Price et al. (2016)	Nunn Creek, CO	3.434	0.980	0.127	9.741	0.580	20.745	1.610	T
Ensign and Doyle (2005)	Snapping Turtle Canal, NC	14.942	0.813	0.338	1.360	0.824	17.000	1.000	T
Ensign and Doyle (2005)	Snapping Turtle Canal, NC	5.358	0.701	0.131	1.401	0.824	17.000	1.000	T
Ensign and Doyle (2005)	Snapping Turtle Canal, NC	9.892	0.718	0.082	20.197	1.319	17.000	1.000	
Ensign and Doyle (2005)	Snapping Turtle Canal, NC	5.049	0.831	0.187	30.502	0.948	17.000	1.000	T
Gooseff (2007b)	Teton Pines Stream, WY	32.896	0.960	0.230	5.939	0.718	25.165	1.041	
Cheong et al. (2007)*est	Shingobee River, MN	23.194	0.476	0.306	14.919	2.091	8.427	1.073	
Cheong et al. (2007)*est	Shingobee River, MN	42.533	0.566	0.041	10.025	2.618	8.860	1.053	
Czernuszenko et al. (1998)*est	Botna river, Moldova	133.632	0.533	0.056	11.575	4.643	20.500	1.120	T
Czernuszenko et al. (1998)*est	Botna river, Moldova	107.393	0.713	0.100	4.464	4.643	20.500	1.120	
Czernuszenko et al. (1998)*est	Botna river, Moldova	72.366	0.875	0.097	4.726	4.643	20.500	1.120	
Czernuszenko et al. (1998)*est	Byk river, Moldova	84.843	0.946	0.139	2.027	6.258	18.214	1.230	
Czernuszenko et al. (1998)*est	Byk river, Moldova	83.065	0.949	0.142	2.293	6.258	18.214	1.230	
Czernuszenko et al. (1998)*est	Kogilnik river, Moldova	10.934	1.237	0.085	39.193	6.089	5.750	1.310	
Godfrey and Frederick (1970)*est	Clinch River B, VA	10.674	0.963	0.058	55.826	7.299	33.152	1.015	

Godfrey and Frederick (1970)*est	Clinch River B, VA	7.638	0.929	0.091	43.071	7.128	29.508	1.011	
Godfrey and Frederick (1970)*est	Clinch River B, VA	6.601	0.769	0.086	41.305	6.192	29.170	1.009	
Godfrey and Frederick (1970)*est	Clinch River B, VA	10.405	0.808	0.140	42.189	6.203	28.921	1.009	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	9.830	0.921	0.202	35.604	5.539	25.034	1.010	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	5.138	1.173	0.128	42.213	8.653	27.519	1.001	
Godfrey and Frederick (1970)*est	Clinch River B, VA	6.151	0.989	0.128	38.067	7.498	27.316	1.005	
Godfrey and Frederick (1970)*est	Clinch River B, VA	11.031	0.876	0.196	46.927	7.518	27.050	1.006	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	10.029	1.029	0.267	40.415	6.723	23.261	1.008	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	7.096	0.974	0.115	31.354	7.712	24.623	1.001	
Godfrey and Frederick (1970)*est	Clinch River B, VA	11.951	0.822	0.185	53.640	7.737	24.369	1.004	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	9.721	1.074	0.291	44.653	7.043	21.141	1.007	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	9.059	0.985	0.319	57.796	6.407	21.346	1.010	
Godfrey and Frederick (1970)*est	Clinch River B, VA	4.976	1.350	0.475	43.514	6.418	21.117	1.012	T
Godfrey and Frederick (1970)*est	Copper Creek A, VA	47.699	0.628	0.246	39.210	2.607	27.826	1.021	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	150.692	0.679	0.553	4.992	2.446	27.920	1.019	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	171.797	0.668	0.115	4.186	2.502	26.912	1.022	T
Godfrey and Frederick (1970)*est	Copper Creek A, VA	90.774	0.658	0.111	11.001	2.381	30.172	1.039	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	171.646	0.706	0.104	4.745	3.059	32.993	1.012	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	78.991	0.764	0.140	11.755	2.921	37.024	1.033	
Godfrey and Frederick (1970)*est	Clinch River B, VA	7.844	0.574	0.052	25.309	3.171	47.734	1.009	
Godfrey and Frederick (1970)*est	Clinch River B, VA	26.298	0.609	0.061	11.332	2.924	46.839	1.009	
Godfrey and Frederick (1970)*est	Clinch River B, VA	25.700	0.823	0.154	36.461	6.184	71.066	1.001	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	14.120	0.638	0.067	22.363	4.294	61.341	1.005	
Godfrey and Frederick (1970)*est	Clinch River B, VA	55.612	0.697	0.079	7.317	3.860	59.335	1.006	
Godfrey and Frederick (1970)*est	Clinch River B, VA	11.879	0.652	0.103	13.930	2.268	37.438	1.008	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	45.025	0.781	0.069	5.525	3.569	54.276	1.004	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	11.436	0.656	0.116	12.145	2.158	34.809	1.007	T
Godfrey and Frederick (1970)*est	Clinch River B, VA	44.673	0.767	0.420	7.473	2.815	49.854	1.007	
Godfrey and Frederick (1970)*est	Clinch River B, VA	9.973	0.655	0.149	14.128	1.848	33.930	1.010	
Godfrey and Frederick (1970)*est	Clinch River B, VA	2.793	0.685	0.141	19.666	1.749	33.774	1.012	
Godfrey and Frederick (1970)*est	Clinch River E, VA	19.345	1.039	0.256	21.079	4.912	71.582	1.017	
Godfrey and Frederick (1970)*est	Clinch River E, VA	50.342	1.523	0.353	11.316	7.563	68.371	1.073	
Godfrey and Frederick (1970)*est	Clinch River E, VA	59.535	1.311	0.377	13.955	6.381	59.155	1.053	
Godfrey and Frederick (1970)*est	Clinch River E, VA	154.730	1.203	0.169	4.716	4.371	56.595	1.045	T
Godfrey and Frederick (1970)*est	Clinch River E, VA	48.232	1.247	0.268	12.242	6.720	63.260	1.105	T
Godfrey and Frederick (1970)*est	Clinch River E, VA	102.912	0.983	0.183	11.281	4.208	53.863	1.050	
Godfrey and Frederick (1970)*est	Clinch River E, VA	202.231	1.183	0.214	4.674	5.084	77.143	1.040	
Godfrey and Frederick (1970)*est	Clinch River E, VA	224.658	1.589	0.656	14.245	9.128	50.000	1.012	
Godfrey and Frederick (1970)*est	Clinch River E, VA	213.723	1.703	0.887	2.208	6.118	48.768	1.019	
Godfrey and Frederick (1970)*est	Clinch River E, VA	95.833	1.833	0.441	2.218	5.423	43.304	1.023	
Godfrey and Frederick (1970)*est	Clinch River E, VA	124.332	1.617	0.449	3.398	6.363	65.395	1.014	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	256.389	0.734	0.199	8.112	5.028	17.479	1.022	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	228.213	0.809	0.347	6.055	4.923	18.624	1.039	T
Godfrey and Frederick (1970)*est	Copper Creek A, VA	249.702	0.737	0.427	4.638	4.578	18.590	1.039	T

Godfrey and Frederick (1970)*est	Copper Creek A, VA	65.973	0.705	0.193	56.166	5.215	23.810	1.007	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	291.092	0.780	0.516	8.169	4.929	23.725	1.012	
Godfrey and Frederick (1970)*est	Copper Creek A, VA	233.666	0.871	0.893	6.977	4.841	25.049	1.033	T
Godfrey and Frederick (1970)*est	Copper Creek A, VA	131.451	1.035	0.293	7.815	4.980	22.991	1.062	T
Godfrey and Frederick (1970)*est	Copper Creek A, VA	215.252	0.793	0.305	4.640	4.609	22.735	1.052	
Godfrey and Frederick (1970)*est	Coachella Canal, CA	4.673	0.903	0.046	51.871	7.569	15.504	1.040	
Godfrey and Frederick (1970)*est	Coachella Canal, CA	6.229	0.916	0.280	46.251	8.044	16.531	1.040	T
Graf (1995)*est	Colorado River, AZ	40.033	0.852	0.037	1.546	2.186	12.427	1.012	T
Graf (1995)*est	Colorado River, AZ	81.774	0.710	0.012	0.047	2.448	11.021	1.174	
RowińSki et al. (2007)*est	Narew River, Poland	213.209	0.977	0.291	3.243	10.539	12.089	1.078	

*est: estimated data using TSM-SC-SAHEL

T: Testing set

국문초록

하천 오염물질 혼합 해석을 위한 저장대 모형의 매개변수 산정법 및 경험식 개발

서울대학교 대학원

건설환경공학부

노 효 섭

하천의 수질을 관리하기 위해서는 자연하천에서 유입된 물질이 이송되고 지체되는 메카니즘을 규명하고 이해하는 것이 필요하다. 하천에서의 물질 혼합을 이해하기 위해 수행된 추적자 실험 연구들에 따르면 자연하천에서 계측되는 농도곡선에서는 가파른 상승부와 긴 꼬리기 관측되는 것으로 알려졌다. 이러한 현상은 주로 물질이 흐르는 본류대와 잠시 물질이 포획되었다가 재방출되는 본류대와 저장대 간의 물질교환 효과 때문에 일어난다고 알려져 있다. 이러한 저장대 물질교환 효과를 모사하는 저장대모형 중 Transient Storage zone Model (TSM)은 가장 광범위하게 이용되는 모형으로, 이를 이용하기 위해선 네 가지의 저장대 매개변수를 보정하여야 한다. 네 가지 저장대 매개변수를 결정하는 방법으로는 일반적으로 현장실험에서 측정된 농도곡선을 이용한 역산모형이 이용된다. 그러나 매개변수가 필요할 때마다 추적자실험을 수행하여 역산모형을 이용하는 것은 현실적으로 불가능한 경우가 있어 이러한 경우에는 비교적

취득하기 쉬운 수리지형학적 인자들을 이용해 매개변수를 산정하는 방법이 이용될 수 있다. 따라서 본 연구에서는 TSM 매개변수를 결정하기 위해 두 가지 방법을 제시하였다. 첫 번째로, 전역 최적화 프레임워크인 Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL)을 이용한 역산모형 기반 TSM 매개변수 산정 프레임워크를 제시하였다. 둘째로는 기호회귀법 라이브러리인 GPTIPS 를 이용한 다중유전자 유전 프로그래밍(Multigene Genetic Programming, MGGP) 과 주성분회귀법(Principal Components Regression, PCR)을 통해 네 가지 매개변수 별로 각 두 개씩의 경험식이 개발되었다. 개발된 경험식들의 성능평가 결과, 선행 연구에서 제시된 저장대 매개변수 식에 비해 본 연구에서 제시된 방법이 대체적으로 우수한 것으로 나타났다. 결과적으로 본 연구에서는 분석을 통해 실무적으로 활용 가능한 TSM 매개변수 산정 프레임워크와 경험식들이 제시되었으며, 이 방법들은 추적자 실험 자료의 유무에 따라 TSM 의 매개변수 결정에 유용하게 사용될 것으로 기대된다.

주요어: 물질혼합, 저장대모형(Transient Storage Model, TSM), 매개변수 산정, 다중유전자 유전 알고리즘(Multigene Genetic Programming, MGGP), 주성분회귀(Principal Components Regression, PCR), Shuffled Complex-Self Adaptive Hybrid EvoLution (SC-SAHEL),

학번: 2017-29424